

Some parts of this thesis may have been removed for copyright restrictions.

If you have discovered material in AURA which is unlawful e.g. breaches copyright, (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please read our [Takedown Policy](#) and [contact the service](#) immediately

Computer-Aided
Categorisation and Quantification of Connectives
in English and Arabic
(Based on Newspaper Text Corpora)

Volume (3)

Adnan Jabbar Radhi Al-Jubouri

Doctor of Philosophy

The University of Aston in Birmingham

July 1987

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the author's prior, written consent.

PREFACE TO VOLUME (3)

This volume is divided into two distinct components. The first contains the last two chapters of the thesis: Chapters 9 and 10. Chapter 9 is a contrastive exposition of the properties of connectives, both quantitative and qualitative, as observed in the English and Arabic corpora. The discussion is related to the description of connectives elaborated in Volume 2 and is accomplished through the method outlined in Volume 1, particularly Chapter 2.

Chapter 10 is a conclusive chapter. Results of the contrastive analysis are here brought together and a number of distinctive generalisations are made. These are then utilised as a basis for making a number of pedagogical implications for EFL teaching of writing, the point we departed from in Chapter 1.

The second component of this Volume consists of 7 appendices that constitute the first of three parts of appendices. Appendix 1 discusses the status of the word as a unit of linguistic measurement as adopted in this project. Particular emphasis is placed on the problematic nature of the concept of the word in Arabic. Appendix 2 reviews in a broad perspective a number of text-based approaches available in the literature; the discussion can serve as an introduction to the theoretical discussion given in Volume 1, particularly in Chapter 2. Appendix 3 is related to Chapter 3 and discusses some criteria for an efficient conception of the term text. The rest of the appendices (App. 4-7) are related to Chapter 5 in Volume 1. The other two parts of the appendices are given in Volume 4.

CONTENTS

	PAGE
Preface	2
Contents	3
List of Tables	10
List of Figures	12
 Chapter 9 <u>Connectives in English and Arabic:</u>	
<u>A Contrastive Account</u>	16
9.0 Perspective	16
9.1 The Quantitative Characterisation of Connectives	18
9.1.1 Introduction	18
9.1.2 Variation in Sentence-Connective Distribution	19
9.1.3 Type-token Indices	22
9.1.4 Frequency Distribution	25
9.1.5 Measure of Repetitiveness	25
9.1.6 Variation in Entropy and Redundancy	30
9.1.7 Variation in Growth	32
9.1.8 Preview of Functional Categories	37
9.2 Interlingual Variations in the Functional Categories	39
9.2.1 Introduction	39
9.2.2 Additive Connectives	40
9.2.2.1 Quantitative Variations of Additive Connectives	40
9.2.2.2 Textual Variations of Additive Connectives	49
9.2.2.2.1 Appending	49

	PAGE
9.2.2.2.2 Enumeration	53
9.2.2.2.3 Amplification	55
9.2.2.2.4 Comment	56
9.2.2.2.5 Continuity	59
9.2.2.2.6 Coupling	62
9.2.3 Comparative Connectives	63
9.2.3.1 Quantitative Variations of Comparative Connectives	63
9.2.3.2 Textual Variations of Comparative Connectives	70
9.2.3.2.1 Similarity	70
9.2.3.2.2 Degree	70
9.2.4 Alternative Connectives	74
9.2.4.1 Quantitative Variations of Alternative Connectives	74
9.2.4.2 Textual Variations of Alternative Connectives	79
9.2.5 Reformulatory Connectives	82
9.2.5.1 Quantitative Variations of Reformulatory Connectives	82
9.2.5.2 Textual Variations of Reformulatory Connectives	91
9.2.5.2.1 Restatement	91
9.2.5.2.2 Exemplification	92
9.2.6 Orientative Connective	94
9.2.6.1 Quantitative Variations of Orientative Connectives	94
9.2.6.2 Textual Variations of Orientative Connectives	102

	PAGE
9.2.6.2.1 Adjustment	102
9.2.6.2.2 Confirmation	106
9.2.7 Temporal Connectives	107
9.2.7.1 Quantitative Variations of Temporal Connectives	107
9.2.7.2 Textual Variations of Temporal Connectives	116
9.2.7.2.1 Temporal Sequence	116
9.2.7.2.2 Simultaneity	125
9.2.7.2.3 Temporal Positioning	130
9.2.7.2.4 Span	134
9.2.7.2.5 Circumstance	136
9.2.8 Spatial Connectives	141
9.2.8.1 Quantitative Variations of Spatial Connectives	141
9.2.8.2 Textual Variations of Spatial Connectives	145
9.2.9 Causal Connectives	147
9.2.9.1 Quantitative Variations of Causal Connectives	147
9.2.9.2 Textual Variations of Causal Connectives	156
9.2.9.2.1 Cause-Reason	156
9.2.9.2.2 Result-Inference	158
9.2.9.2.3 Condition	162
9.2.9.2.4 Degree-Magnitude	164
9.2.9.2.5 Purpose	166
9.2.10 Adversative Connectives	167
9.2.10.1 Quantitative Variations of Adversative Connectives	167

	PAGE
9.2.10.2 Textual Variations of Adversative Connectives	175
9.2.10.2.1 Antitheticity	175
9.2.10.2.2 Contrast	178
9.3 Range of Operationality	180
9.3.1 Preliminaries	180
9.3.2 Types of Range	181
9.3.3 Range of Connectives: A General Profile	183
9.3.4 Range of Additive Connectives	187
9.3.5 Range of Alternative Connectives	189
9.3.6 Range of Comparative Connectives	191
9.3.7 Range of Reformulatory Connectives	192
9.3.8 Range of Orientative Connectives	194
9.3.9 Range of Temporal Connectives	196
9.3.10 Range of Spatial Connectives	197
9.3.11 Range of Causal Connectives	198
9.3.12 Range of Adversative Connectives	200
9.4 Conclusion	202
Footnotes to Chapter 9	205
Chapter 10 <u>Conclusions and Pedagogical Implications</u>	206
10.0 Perspective	206
10.1 Text Organisation	207
10.2 Text Organisation in English and Arabic	209
10.2.1 Preliminaries	209
10.2.2 Textual Grouping in English and Arabic	210
10.2.3 Textual Sequencing in English and Arabic	211

	PAGE
10.3 Textual Organisation and Connectivity	219
10.3.1 Preliminaries	219
10.3.2 Types of Connectivity	219
10.3.3 Extension	220
10.3.3.1 Operational Scope	220
10.3.3.2 Types of Extension	221
10.3.3.3 Extension via Additivity	221
10.3.3.3.1 Textual Role of Additivity	221
10.3.3.3.2 Additivity in English and Arabic	222
10.3.3.4 Extension via Comparison	232
10.3.3.4.1 Textual Role of Comparison	232
10.3.3.4.2 Comparison in English and Arabic	233
10.3.4 Variation	235
10.3.4.1 Operational Scope	235
10.3.4.2 Types of Variation	236
10.3.4.3 Variation via Adversativity	236
10.3.4.3.1 Textual Role of Adversativity	236
10.3.4.3.2 Adversativity in English and Arabic	237
10.3.4.4 Variation via Alternation	239
10.3.4.5.1 Textual Role of Alternation	239
10.3.4.5.2 Alternation in English and Arabic	240
10.3.5 Elaboration	241
10.3.5.1 Operational Scope	241
10.3.5.2 Types of Elaboration	242
10.3.5.3 Elaboration via Reformulation	242
10.3.5.3.1 Textual Role of Reformulation	242

	PAGE
10.3.5.3.2 Reformulation in English and Arabic	243
10.3.5.4 Elaboration via Orientation	245
10.3.5.4.1 Textual Role of Orientation	245
10.3.5.4.2 Orientation in English and Arabic	246
10.3.6 Enhancement	248
10.3.6.1 Operational Scope	248
10.3.6.2 Types of Enhancement	248
10.3.6.3 Enhancement via Temporality	249
10.3.6.3.1 Textual Role of Temporality	249
10.3.6.3.2 Temporality in English and Arabic	250
10.3.6.4 Enhancement via Spatial Connectives	254
10.3.6.4.1 Textual Role of Spatial Connectives	254
10.3.6.4.2 Spatial Connectives in English and Arabic	254
10.3.6.5 Enhancement via Causality	255
10.3.6.5.1 Textual Role of Causality	255
10.3.6.5.2 Causality in English and Arabic	256
10.3.7 Concluding Remarks	259
10.4 Implications for Teaching EFL Writing Skills at the Undergraduate Level	261
10.4.1 Preliminaries	261
10.4.2 The Writing Programme: An Introduction	262
10.4.3 Some Guidelines for Designing a Writing Programme	264
10.4.3.1 Requirements	264
10.4.3.2 Some Pedagogical Suggestions	265
10.4.4 Typology of Exercises	272
10.4.5 Concluding Remarks	277

	PAGE
10.5 Suggestions for Further Research	278
10.5.1 Text Analysis	278
10.5.2 Computerised Corpus Analysis	281
10.5.3 Lexicography	282
10.5.4 EFL Writing Pedagogy	283
10.6 A Final Note	284
Footnotes to Chapter 10	286

Appendices

Part 1 <u>Non-Statistical Appendices</u>	288
Appendix 1 Reflections on the Word as a Unit of Linguistic Measurement	288
Appendix 2 Approaches to Text-Based Studies	340
Appendix 3 Toward a Definition of Text: Criteria and Requirements	357
Appendix 4 Manual Input: Problems of Transliteration	359
Appendix 5A OCP Command File for Concordance Generation of English	376
Appendix 5B OCP Command File for Concordance Generation of Arabic	377
Appendix 6 Steps for Accessing OCP	378
Appendix 7A A Sample Page from the English Corpus Containing Tagged Connectives	383
Appendix 7B A Sample Page with Deleted Connectives	384
Appendix 7C A Sample Page with Text Deleted and Connectives Retained	385

LIST OF TABLES

	PAGE
9.1 Levin's measures of compactness of the distribution of connectives in the English and Arabic corpora	30
9.2 Entropy and redundancy of connectives in the English and Arabic corpora	31
9.3 Frequency distribution of connectives of temporal sequencing occurring within a distance of 20 words in the Arabic corpus	124
9.4 Frequency distribution of connectives of temporal sequencing occurring within a distance of 20 words in the English corpus	125
9.5 Connective cores used for constructing time relators with referential elements (as observed in the English corpus)	132
9.6 Connective cores used for constructing time relators with referential elements (as observed in the Arabic corpus)	132
9.7 Distribution of operational range of connectives	184
9.8 Distribution of range of additive connectives	188
9.9 Distribution of range of comparative connectives	190
9.10 Distribution of range of alternative connectives	191

	PAGE
9.11 Distribution of range of reformulatory connectives	193
9.12 Distribution of range of orientative connectives	195
9.13 Distribution of range of temporal connectives	196
9.14 Distribution of range of spatial connectives	197
9.15 Distribution of range of causal connectives	198
9.16 Distribution of range of adversative connectives	200

LIST OF FIGURES

	PAGE
9.1 Comparison of the distributions of connectives per sentence in the two corpora	20
9.2 Tokens and Types of Connectives in the English and Arabic corpora	20
9.3 Comparison of the distribution of connectives in the English and Arabic corpora	26
9.4 Comparison of the distribution of connective types in the English and Arabic corpora	26
9.5 Comparison of the distribution of the 10 most frequent connectives	27
9.6 Comparison of growth of connective tokens	33
9.7 Comparison of growth of connective types	33
9.8 Comparison of growth of connective types within tokens in the two corpora	36
9.9 Comparison of the distribution of categories of connectives	36
9.10 Comparison of shares of connective categories in connective mass in the English and Arabic corpora	38
9.11 Comparison of the distribution of connective types within categories of connectives	38
9.12 Comparison of the distribution of additive connectives	41
9.13 Comparison of the distribution of additive connectives within connective tokens	41
9.14 Comparison of "global" growth of additive connectives	44

	PAGE
9.15 Comparison of "local" growth of additive connectives	44
9.16 Comparison of the distribution of categories of additive connectives in the English and Arabic corpora	46
9.17 Comparison of distribution of shares of additive categories	46
9.18 Comparison of the distribution of comparative connectives	64
9.19 Comparison of the distribution of comparative connectives within connective tokens	64
9.20 Comparison of the global growth of comparative connectives	67
9.21 Comparison of the local growth of comparative connectives	67
9.22 Comparison of the distribution of categories of comparative connectives in the two corpora	69
9.23 Comparison of the distribution of shares of comparative categories	69
9.24 Comparison of the distribution of alternative connectives	76
9.25 Comparison of the distribution of alternative connectives within connective tokens	76
9.26 Comparison of the global growth of alternative connectives	78
9.27 Comparison of the local growth of alternative connectives	78
9.28 The semantic content of propositions entering in an alternative relation in (1) English (2) Arabic	81
9.29 Comparison of the distribution of reformulatory connectives	83
9.30 Comparison of the distribution of reformulatory connectives within connective tokens	83
9.31 Comparison of the global growth of reformulatory connectives	86

9.32	Comparison of local growth of reformulatory connectives	86
9.33	Comparison of the distribution of the categories of reformulatory connectives in the two corpora	89
9.34	Comparison of the distribution of shares of reformulatory connectives	89
9.35	Comparison of the distribution of orientative connectives	95
9.36	Comparison of the distribution of orientative connectives within connective tokens	95
9.37	Comparison of the global growth of orientative connectives	98
9.38	Comparison of the local growth of orientative connectives	98
9.39	Comparison of the distribution of categories of orientative connectives in the two corpora	100
9.40	Comparison of the distribution of shares of orientative categories	100
9.41	Comparison of the distribution of temporal connectives	108
9.42	Comparison of the distribution of temporal connectives within connective tokens	108
9.43	Comparison of the global growth of temporal connectives	110
9.44	Comparison of the local growth of temporal connectives	110
9.45	Comparison of the distribution of the categories of temporal connectives in the two corpora	113
9.46	Comparison of the distribution of shares of temporal categories	113
9.47	Comparison of the functional scope of "until" vs. "'ilā 'an" and "ḥattā"	135
9.48	Comparison of the distribution of spatial connectives	142

	PAGE
9.49 Comparison of the distribution of spatial connectives within connective tokens	142
9.50 Comparison of the global growth of spatial connectives	144
9.51 Comparison of the local growth of spatial connectives	144
9.52 Comparison of the distribution of causal connectives	148
9.53 Comparison of the distribution of causal connectives within connective tokens	148
9.54 Comparison of global growth of causal connectives	150
9.55 Comparison of local growth of causal connectives	150
9.56 Comparison of the distribution of categories of causal connectives in the two corpora	153
9.57 Comparison of the distribution of shares of causal categories	153
9.58 Clustering of connectives introducing result/consequence on the relation continuum	159
9.59 Comparison of the distribution of adversative connectives	168
9.60 Comparison of the distribution of adversatives within connective tokens	168
9.61 Comparison of global growth of adversative connectives	171
9.62 Comparison of local growth of adversative connectives	171
9.63 Comparison of the distribution of the categories of adversative connectives in the two corpora	174
9.64 Comparison of the distribution of shares of adversative categories	174
10.1 Text development via appending in Arabic	224

Connectives in English and Arabic: A Contrastive Account

9.0 Perspective

One of the main tasks that this project has set out to achieve is the provision of a textual contrastive analysis of the textual functioning and rhetorical patterns associated with connectives in English and Arabic. We have stated (see Ch. 2) that the investigatory apparatus adopted aims first at proffering an adequate description of the phenomenon before any analytic effort is expended on accomplishing the contrastive task. The description falls into two parts: textual, which was provided in Ch. 6, and quantitative, introduced in Ch. 7 and 8.

The aim of this chapter is to reconsider the description within a contrastive perspective and make a detailed and useful comparison of textual and quantitative variations in the behaviour of connectives in each corpus. More specifically, we would like to achieve the following tasks:

1. Give a general contrastive account of the various functional categories of connectives.
2. Provide a contrastive statistical account of each main functional category of connectives.
3. Compare the size and nature of the distribution of each subcategory and monitor the behaviour of the top most frequent connectives.

4. Offer a detailed functional/rhetorical comparison of the role of connectives within each subcategory, pointing out their unique and shared textual properties, the constraints they operate under and their patterns of divergence.

5. Examine connective range of operationality in both corpora.

In order to systematise the contrastive account, the quantitative profile is given first as a background against which the textual variations are to be identified and examined. Accordingly, the content of the chapter is organised into four main sections. The first investigates variations in the main features that constitute the general statistical profile of connectives. Each feature is closely examined and the indices and measurements are set next to each other and closely compared.

Section 2 is the longest and constitutes the main bulk of the chapter. It is a detailed examination of the variations, both quantitative and textual, in the behaviour of each functional category of connectives in the two corpora.

Section 3 is an examination of the operational range that connectives impose on text constituents. This is examined in two perspectives. The first is general and deals with global tendencies; the second is specific and deals with range as imposed by various categories.

Finally section 4 offers brief conclusive remarks. These remarks are later picked up (see Ch. 10), reconsidered and related to the whole study.

9.1 The Quantitative Characterisation of Connectives

9.1.1 Introduction

The statistical accounts of connectives that have been generated assist to a considerable extent in proffering a characterisation of the textual organisation in each language and lead to a better understanding of the nature of connectivity. A sharper insight is developed when the main quantitative aspects of connectives are related to each other and a differential statement is formulated.

This is what this section intends to achieve. The general aim is the expression of the overall difference in the numerical features of connectives in the two corpora and the provision of an interpretation that can later be utilised in pointing out the nature of the textual role of connectives. This aim is achieved via the implementation of a number of tasks:

1. Comparison of the quantitative sentence-connective relation in the two corpora.
2. Consideration of the relative frequency and the various indices that have been obtained by type-token mathematics.
3. Description of the variations in patterns of repetitiveness of connectives in the two corpora. This includes comparison of repeat rate and gap/interval distributions.
4. Inspection of growth in the two corpora.
5. Interpretation of the informational indices of entropy and redundancy in the two corpora.

9.1.2 Variation in Sentence-Connective Distribution

Our starting point is the striking difference in the connective-sentence distribution, which is displayed in Figure (9.1). This can be summarised as follows:

1. Although the corpora are comparable in size, the number of sentences and connectives are widely different and represent two opposing angles: a) On the one hand, the number of connectives in the Arabic corpus is far greater than that of the English corpus: 16,995 connective tokens in Arabic vs. 9,596 in English; this difference is equivalent to 77%. b) On the other hand, the number of sentences in the Arabic corpus is 31% smaller than its equivalent in the English corpus (8,060 sentences in Arabic vs. 11,671 in English). Thus, while the mean connectives per sentence in Arabic is 2.1086, it is 0.8222 in English, a difference equivalent to approximately 157%.

2. The distribution of connectives per sentence points to a sharp contrast in the two languages: a) While sentences that do not include connectives represent 46% of total sentences in English, they are equivalent to only 11.6% of sentences in the Arabic corpus. In other words, more than 88% of sentences in Arabic incorporate connectives in contrast to 54% of English sentences. b) The relative frequency of sentences comprising one connective each is similar in the two corpora: 34% in English vs. 32.3% in Arabic. This means that the difference in the percentages that are calculated in (a) represents, in Arabic, a bigger relative frequency of sentences with two or more connectives than in English. Indeed, sentences that

Fig. 9.1 Comparison of the Distributions of Connectives per Sentence in the Two Corpora

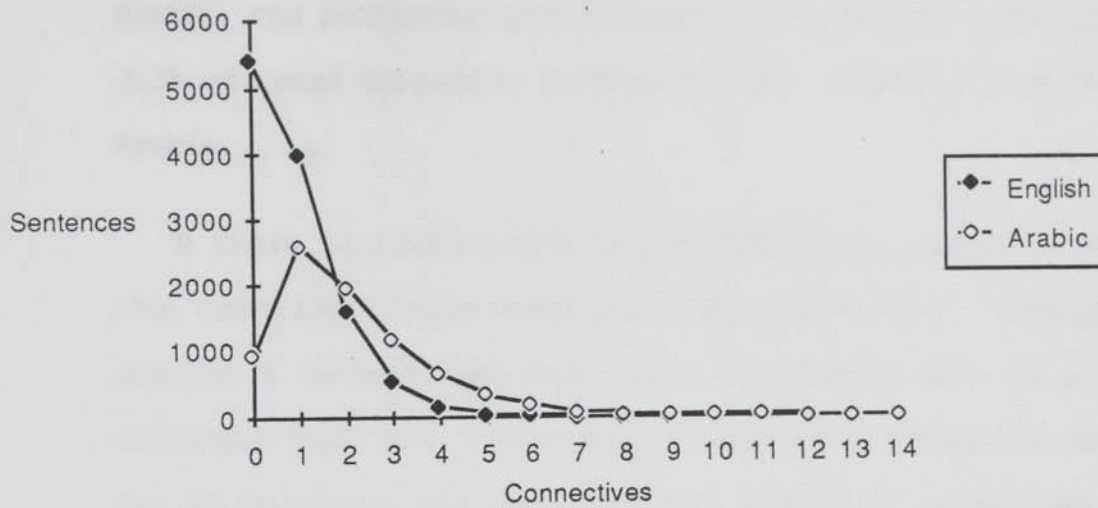
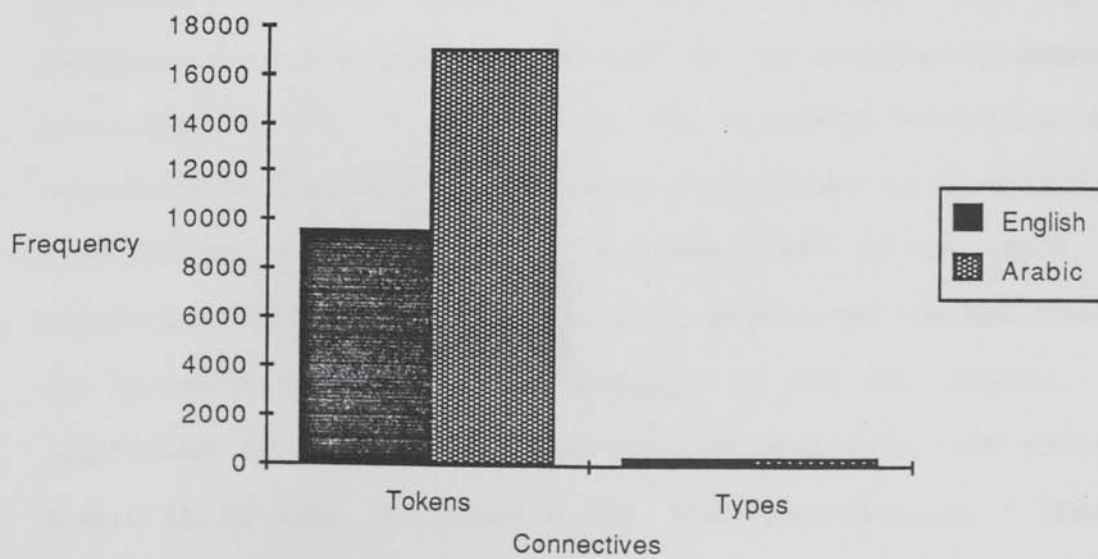


Fig. 9.2 Tokens and Types of Connectives in the English and Arabic Corpora



contain 2 connectives each correspond to 13.5% of total sentences in English but 24.6% in Arabic; sentences having 3 connectives each are equivalent to 4.3% of total sentences in English, but 14.8% in Arabic; and sentences that include 4 connectives each constitute 1.3% of total sentences in English, but slightly less than 9% in Arabic.

3) There is a noticeable difference in the number of sentences that comprise a large number of connectives each. Sentences that contain 8 connectives each is 2 in English but 40 in Arabic; sentences that have 9 connectives each are non-existent in English but 29 in Arabic; and sentences that contain 10 connectives each are one in English but 8 in Arabic. In addition there are sentences in Arabic that contain 11 connectives (8 sentences), 12 connectives (5), 13 connectives (3) and 14 connectives (3). Such sentences are non-existent in the English corpus.

4) There is a noticeable difference in the coefficient of variation in the two corpora. This measure represents the standard deviation of a distribution divided by the arithmetic mean; if multiplied by 100, it expresses the standard deviation as a percentage of the mean. Since as a pure number it constitutes a dimensionless measure, i.e. it is independent of the scale, the coefficient of variation enables us to compare the variabilities of the sentence-connective distributions of the two corpora. The coefficient of variation in the English corpus is 119% while in Arabic it is 81%. In other words, the distribution in English displays a greater variability than its counterpart in the Arabic

corpus does.

To complement this comparison, we tested the mean of connective per sentence in the two corpora for significance by using two-tailed Z-test. The question that is asked is: if we claim that the two means differ, shall we have at least a 99.9% chance of being correct? The null hypothesis is that there is no difference between the two means. The alternative hypothesis specifies that a difference does exist between the two means and this difference can be established at the 0.001 level. The results of the test prove the alternative hypothesis.

9.1.3 Type-token Indices

The next type of quantitative variations that we have observed concerns a number of indices based on the type-token relations. Since the types and tokens differ to a considerable extent in the two corpora, as Figure (9.2) displays, it follows that the measures that make use of these two figures will render an entirely different set of indices.

1. Type-token ratio

The connective type-token ratio differs significantly in the two corpora. Since the English corpus comprises a lower number of tokens (9,596 in English vs. 16,995 in Arabic) but a higher number of types (311 vs. 297), the type-token ratio in English is higher than its counterpart in Arabic (0.0324 vs. 0.0175).

2. Predictability

According to the indices in (1) above, connective types are more

predictable in the Arabic than English corpus. The predictability index in Arabic is 0.9825 while in English it is 0.9676.

3. Concentration

Concentration of connective types is higher in the English than in the Arabic corpus. The index is 0.0236 in English but 0.0130 in Arabic.

4. Token Occurrence Rate

Since the number of connective tokens is higher in the Arabic than English corpus, it follows that the occurrence rate of connective tokens is higher in Arabic than in English. In Arabic one connective is encountered in each sequence of 15 running words, while in English one connective occurs in every sequence of 27 running words.

5. Type Occurrence Rate

This index is 30.855 in the English corpus, which means that a new connective type appears at the rate of 31 successive connective tokens. In the Arabic corpus this index is 57, indicating a lower type repeat rate: a new type is encountered in every 57 successive connective tokens.

6. Hapax and Non-hapax Indices

The number of hapax legomena in the English corpus is slightly higher than in the Arabic corpus: 85 vs. 77. The hapax probability of occurrence in the English corpus is 0.273 while its Arabic counterpart is 0.258. These figures mean that the non-hapax probability of occurrence in the English corpus is slightly lower than its Arabic equivalent: 0.727 vs. 0.742.

7. Exclusivity

The exclusivity index indicates the size of the share that hapaxes reserve in the token mass. It is higher in the English than in the Arabic corpus: 8.86×10^{-3} vs. 4.53×10^{-3} .

8. Variegation

The variegation index is indicative of connective diversification. It is higher in the English corpus, calculated at 27.33%, than in the Arabic, 25.84%.

9. Gravity

This index refers to the occurrence rate of hapaxes within each corpus. It is approximately 113 in the English corpus, but 221 in Arabic. This indicates a higher rate and therefore a bigger gravity factor in English in comparison to Arabic.

10. Stereotypicality

Connectives in the Arabic corpus are more stereotypical than those in the English corpus. The index of stereotypicality, referring to the repeatedness of non-hapaxes, is 76.55 in Arabic. In contrast the index is 42.08 in English.

11. Rhythmicality

This index indicates the extent of repeatedness of connectives in the two corpora. It is higher in the Arabic corpus than in the English: 75.55 vs. 48.08.

12. Consolidation factor

This factor, which refers to the intensity of use of types,

particularly non-hapaxes, is slightly higher in the Arabic than in the English corpus. The index is calculated at 0.742 in Arabic and 0.727 in English.

9.1.4 Frequency Distribution

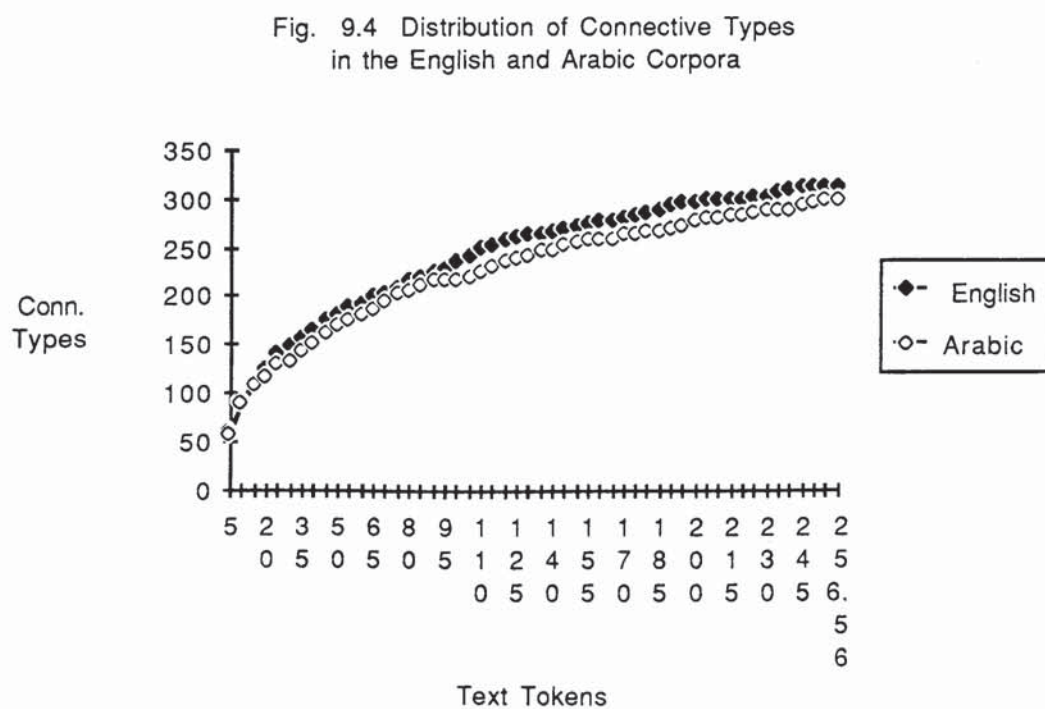
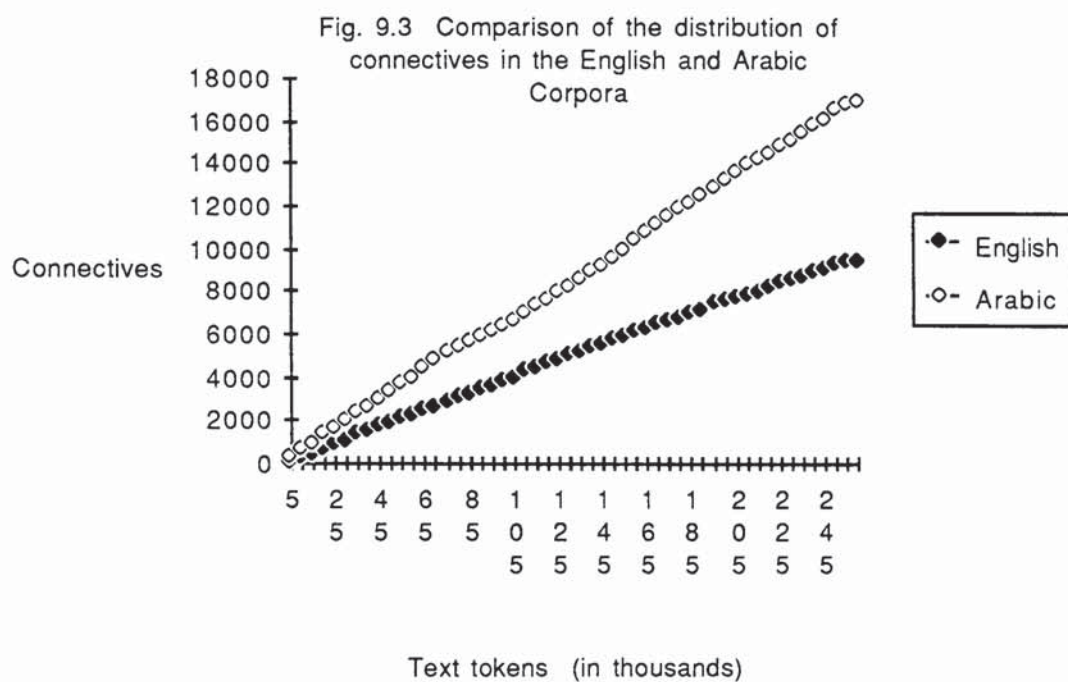
The variation in the frequency distribution of tokens and types in the two corpora are displayed in Figures (9.3-4). The rank distribution in Arabic shows that the top two connectives achieve more than 59% of token coverage. This large share in token mass is achieved by the combined frequencies of the top 10 connectives in the English corpus. The top 10 connectives in Arabic represent collectively 72% of connective token mass. Figure 9.5 compares the frequencies of the first 10 connectives in both corpora. Further, a 75% token coverage is attained by 23 connectives in English and by 13 in Arabic. This is indicative of the reliance of connectivity in Arabic on a relatively small set of connectives.

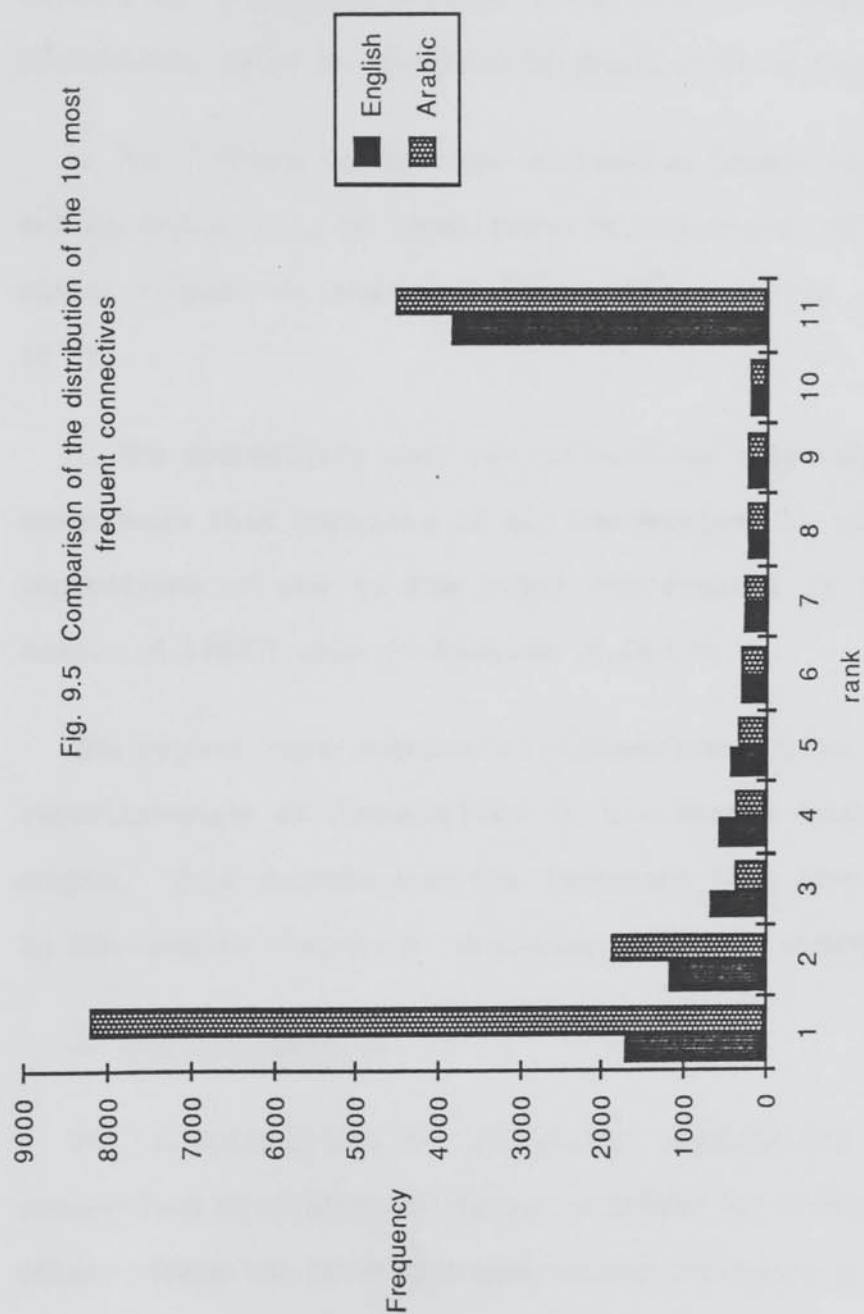
9.1.5 Measures of Repetitiveness

Some measures of repetitiveness have already been mentioned (e.g. token and type occurrence rate, stereotypicality, and rhythmicity). We discuss here three: repeat rate, gap distribution and Levin's interval distribution.

1. Repeat rate

The repeat rate index, calculated as the sum of the squared relative frequencies of connectives, depends upon the frequency distribution of connectives in each corpus and has been used in quantitative linguistic studies as a valid characteristic of





1. and 2. but 3. if 4. as 5. when 6. also 7. because
8. even 9. then 10. so 11. Other Connectives
11. Other Connective Types

1. wa 2. fa 3. kama 4. 'ida 5. lakinna 6. li'anna 7. li
8. hatta 9. bal 10. 'aw 11. Other Connectives
11. Other Connective Types

repetitiveness. Three types of repeat rate have been compared:

a. The probability that two successive random selections from the whole corpus will turn out to be connectives is higher in Arabic (4.39×10^{-3}) than in English (1.39×10^{-3}). In other words, this probability is 44 in 10,000 in Arabic but 14 in 10,000 in English.

b. The probability that two successive random selections from the entire corpus will be repetitions of one or the other connective is, again, higher in Arabic (1.093×10^{-3}) than in English (0.088×10^{-3}).

c. The probability that two connectives selected at random from a great mass that consists of all connectives in the corpus will be repetitions of one or the other connective, is again, higher in Arabic (0.24897) than in English (0.06309).

The repeat rate indices of connectives, thus, display higher repetitiveness of connectives in the Arabic than in the English corpus. This demonstrates the important role that connectives play in the textual/rhetorical structure of Arabic compared to English.

2. Gap Distribution

The distribution of distances separating repetitions of connectives display significant differences when compared to each other. These can be summarised in the following points:

a. The average gap that separates two connectives in the Arabic corpus is lower than that its counterpart in the English corpus. In Arabic it is calculated at 14 words and in English at approximately

26 words. These two figures have been tested for significance and the difference is found to be highly significant.

b. The most frequent gap in both distributions is 0 words, i.e. no separating words occurring between two successive connectives. However, in the Arabic corpus this gap occurs with a frequency of 1,877, which constitutes 11% of the number of gaps. In comparison, this gap occurs in the English corpus with a frequency of 618, which corresponds to 6.4% of the number of gaps.

c. The next most frequent gap is shorter and more frequent in Arabic than in English. In Arabic, it is a gap of 6 words which occurs with a frequency of 749, representing 4.45 of the number of gaps. In English, it is a gap of 8 words which has a frequency of 288, representing 3% of the number of gaps in the corpus.

d. In general, the shorter gaps (0-9 words) are more frequent in the Arabic than in the English corpus. In Arabic, they have a combined frequency of 7,951, which is equivalent to approximately 47% of the total number of gaps in the corpus. In English, the combined frequency of these gaps is 2,962, which corresponds to approximately 31% of the total number of gaps.

e. The longest gap has a frequency of one in each corpus. However, it is shorter in Arabic (245 words) than in English (315 words).

3. Levin's Distribution of Interval

The indices that we have obtained by using Levin's formulae to calculate compactness of the distribution of connectives in both

corpora manifest that although in comparison they are not sharply in discord, those of the English corpus are, nevertheless, consistently higher than their equivalent in Arabic. These indices are given in Table (9.1) below. What the comparative differences indicate is more smoothness in the distribution of connectives in English than

Index -----	English -----	Arabic -----
L-compactness:	0.04967	0.04632
Q-compactness:	0.00012	0.00006

Table 9.1 Levin's measures of Compactness of the Distribution of Connectives in the English and Arabic corpora

in Arabic. This refers to the manner of repetitiveness of connectives, computed independently of their frequency of distribution, and can be considered as a textual characteristics of connectivity.

9.1.6 Variation in Entropy and Redundancy

The computation of the entropy and redundancy indices is based on the rank distribution of connectives in the corpora. A summary of the results is given in Table (9.2) below. (A fuller account is given in Appendices 18 and 19). According to information theory, maximum uncertainty would result in a maximum of "information" being carried by each symbol (connective) in the corpus. When each symbol is carrying a maximum of information, the code would, of course, be utilised to capacity. The unit of measurement of the capacity of the code is the bit. If the distribution of connectives was

Index -----	English -----	Arabic -----
Maximum possible entropy (H')	13.22820	14.05280
Entropy (H)	5.43790	3.98587
Relative entropy (h)	41.10830%	28.36350%
Redundancy	58.89170%	71.63650%

Table 9.2 Entropy and redundancy of connectives in the English and Arabic corpora

equiprobable, the capacity of a set of 9,596 connectives (English) would be 13.22820 bits, and the capacity of a set of 16,995 connectives (Arabic) would be 14.05280. The maximum entropy (H') of Arabic connectives is higher than that of the English and is explainable by the bigger number of ranks in the distribution.

However, the distribution of connectives is far from being equiprobable and therefore the entropy H of the set of connectives is never as great as the maximum entropy H' . The results of the computation shows that the entropy of Arabic connectives is smaller than its counterpart in English: 3.98587 bits vs. 5.43790 bits. Relative entropy (H/H') of Arabic connectives is 0.283635 (given as a percentage in the table 28.3635%). In contrast, relative entropy of English connectives is distinctly higher: 0.411083 (expressed as a percentage in the table: 41.1083%).

The calculations made up enable us to compute the size of redundancy of connectives in each corpus. Redundancy is an essential property of language which permits one to understand what is said or written despite considerable amount of noise. Since it is computed as the complement to unity of the relative entropy, redundancy of connectives is greater in the Arabic than in the

English corpus. In Arabic, it is 0.716365 (given in the table as a percentage 71.6365%) while in English it is calculated as 0.588917 (expressed as a percentage 58.8917%). The interpretation of these figures will be discussed further in the next chapter.

9.1.7 Variation in Growth

Computation of growth of connective tokens and types manifest important variations across the two corpora. We shall consider two types of growth: "global" and "local".

A. "Global" Growth

Comparison of the global growth of connective tokens and types across the corpora lead us to make the following observations.

1. Growth of tokens

a. Due to the significantly bigger number of connective tokens in Arabic than in English, each text interval in Arabic displays a more intensive growth than the counterpart does in English (see Figure 9.6). For instance, the first interval in the Arabic corpus contains 369 connective tokens, while the same interval in the English corpus contains 173. Further, the lowest level of growth within the intervals in Arabic - which is 220 tokens - is larger than the highest level of growth in English - which is 219. The average growth is 331 connective tokens per interval in Arabic and 187 in English.

b. Extrapolation of a million word Arabic corpus is expected to contain 63,287 connective tokens. An English corpus of a similar

Fig. 9.6 Comparison of growth of connective tokens

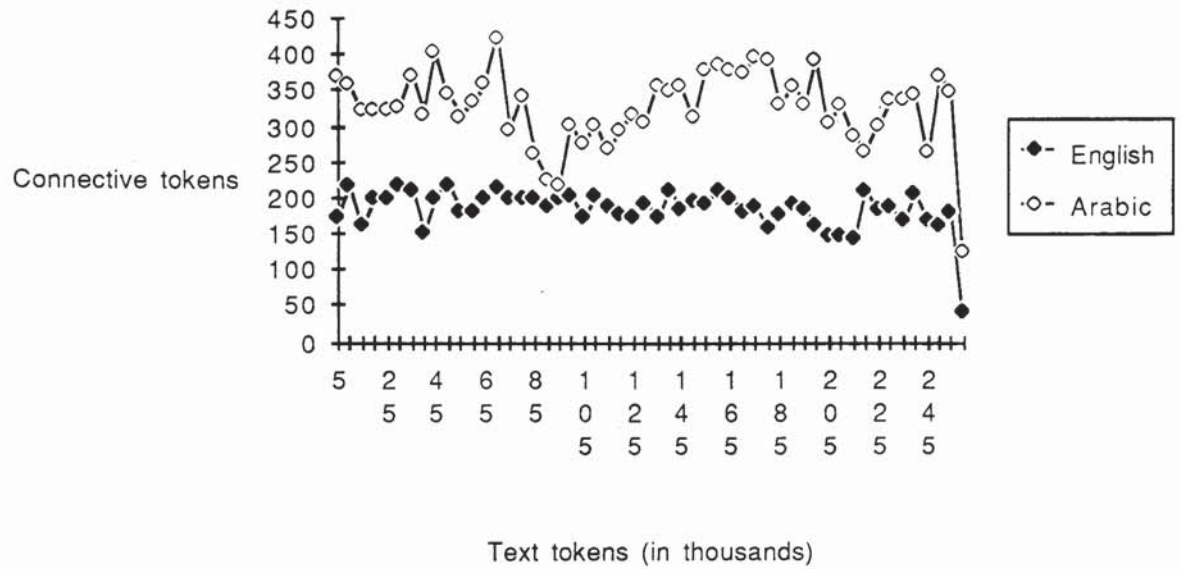
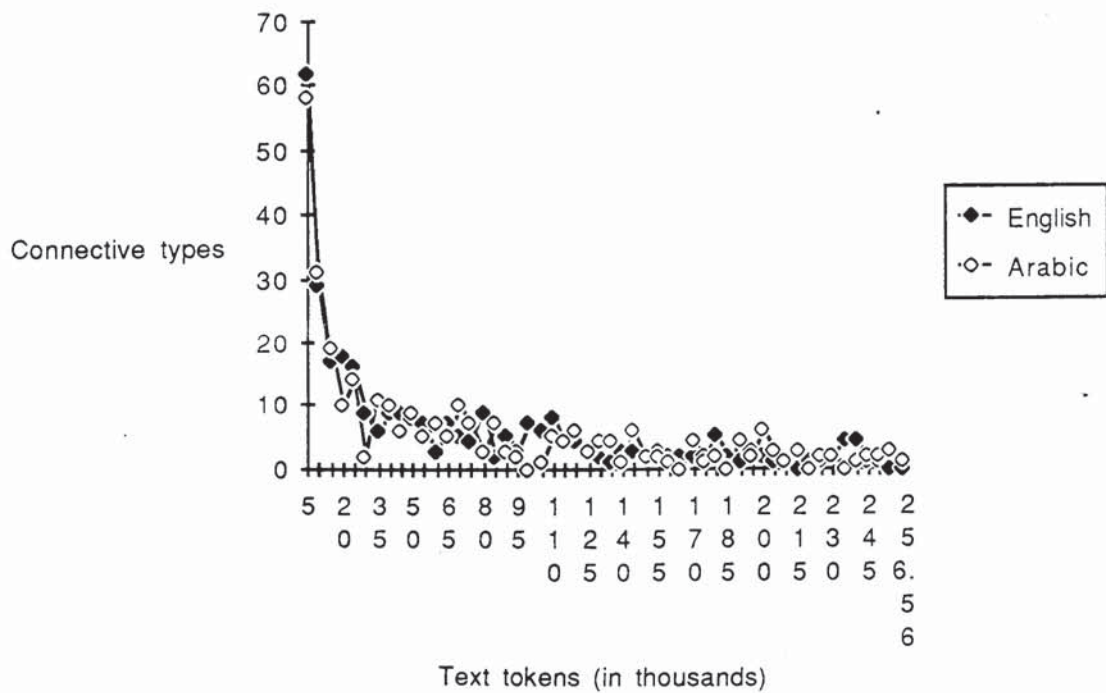


Fig. 9.7 Comparison of growth of connective types



size is expected to have 38,747. The difference in these two figures is a clear indication that Arabic makes a more intensive use of connectives for sustaining cohesion and achieving rhetorical organisation.

2. Growth of Types

a. Due to the similar number of types in both corpora, it has been observed that their growth is similar within the early text intervals (see Figure 9.7). However, inspection of growth in the latter intervals reveals a faster type saturation in English than in Arabic. For instance, the 90th type percentile occurs within intervals 34 in English and 38 in Arabic.

b. Extrapolation of a million word English corpus is expected to contain 541 connective types. An Arabic corpus of a similar size is expected to contain 5.05 types.

c. One interesting calculation derived from the numbers in 2 above is the expected connective TTR. In a million word English corpus the expected index is 0.014, significantly higher than its expected counterpart in an Arabic corpus of a similar size which is 0.008. This is an indication that English makes a richer and more extensive utilisation of its connectives, while, in contrast, Arabic uses connectives more intensively. This factor is central to text organisation in the two languages.

d. The manner and rate of growth of connective types is probably better monitored and compared within intervals of text containing equal numbers of connective tokens, regardless of their size in

terms of text tokens (words). This is investigated under "local" growth.

B. "Local" Growth

The growth of connective types is monitored within intervals of 500 tokens in each corpus and the results have already been discussed in Chapter 7 and given in Tables (7.14) and (7.20). Comparison of growth in the two corpora, which is displayed in Figure (9.8), leads us to make the following observations.

1. The first interval (500 tokens) exhibits a fast growth of types in English (104) in comparison with Arabic (63).

2. However this fast growth in English connective types starts to drop immediately, and the next interval displays similar size of growth in both corpora: 40 types in English as opposed to 43 in Arabic.

3. Despite some fluctuation, growth remains comparatively faster within the next intervals in English, than in Arabic, but the rate drops to a considerable degree. The later intervals in the table show similar growth.

4. Growth of connectives in the Arabic corpus show that more types are encountered even at the later stages. This consistent appearance of types has an important bearing on the expected numbers of types at extrapolated intervals. For instance, using Tuldava's formula, extrapolation to 20,000 tokens yields an expected number of 344 connective types in English and 323 in Arabic. Extrapolation to

Fig. 9.8 Comparison of Growth of Connective Types within Tokens in the two Corpora

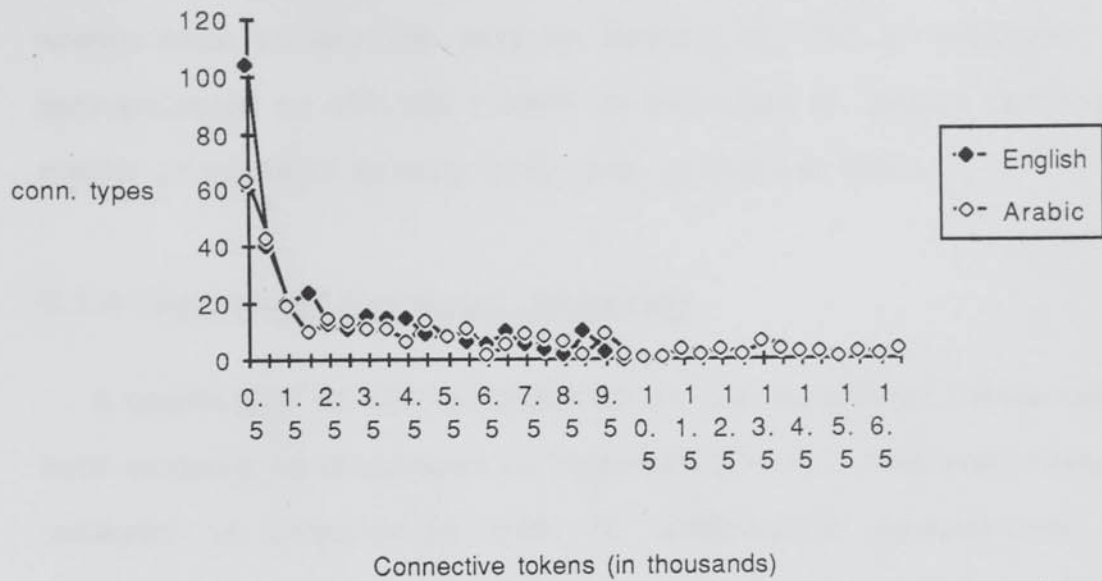
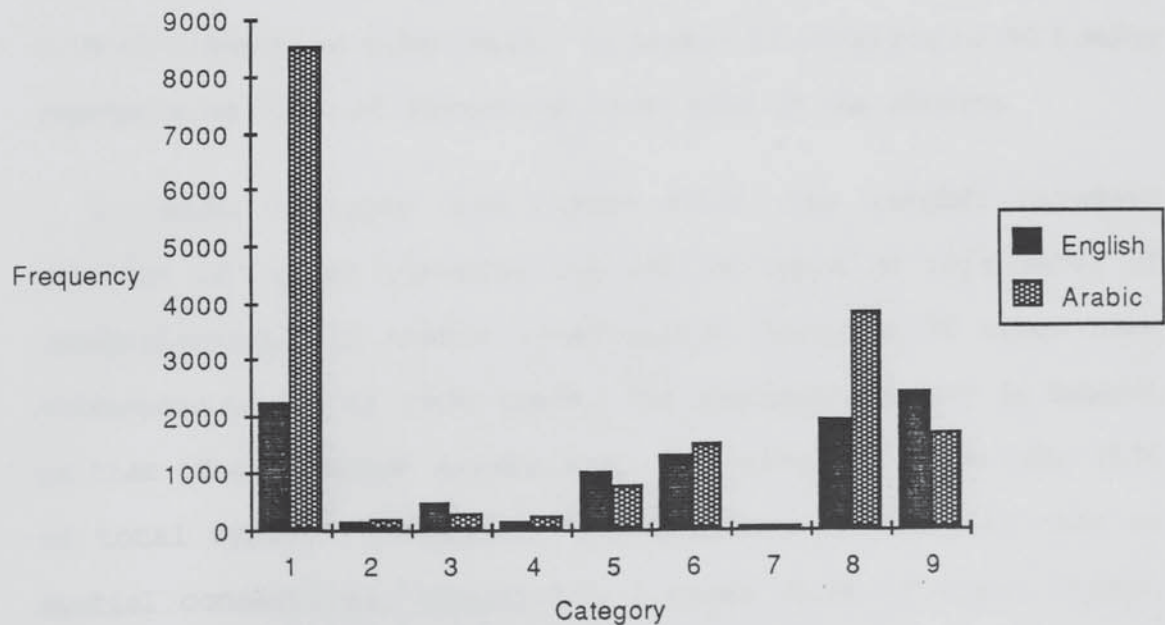


Fig. 9.9 Comparison of the Distribution of Categories of Connectives



1. Additive 2. Alternative 3. Comparative 4. Reformulatory
5. Orientative 6. Temporal 7. Spatial 8. Causal 9. Adversative

50,000 connective tokens is expected to comprise more types in Arabic than in English (431 in Arabic vs. 423 in English). And extrapolation to 100,000 tokens is expected to give a far bigger number of types in Arabic (526) than in English (451).

9.1.8 Preview of Functional Categories

A comparison of the distribution of the functional categories in both corpora is displayed in Figures (9.9-10). The most frequent category in English is that of adversative connectives. It comprises 2,404 tokens that represent 25% of token mass in the corpus. In Arabic the most frequent category is that of additives. It comprises 8,552 tokens representing more than 50% of connective token mass. The smallest category in both corpora is spatial connectives. In English it consists of 50 tokens corresponding to 0.5% of connective token mass. In Arabic it consists of 40 tokens representing 0.24% of connective token mass in the corpora.

In terms of types (see Figure 9.11), the largest category contains 127 types representing 34% of types at this level of categorisation. In Arabic orientatives comprise 92 types that correspond to 24% of total types. The smallest category in English is that of alternative connectives. It contains 2 types only (0.5% of total types). In Arabic, the smallest category is that of spatial connectives, comprising 3 types (0.8% of total types). Fuller comparison of individual categories across the two corpora is given next.

Fig. 9.10 A Comparison of Shares of Connective Categories in Connective Mass in the English and Arabic Corpora

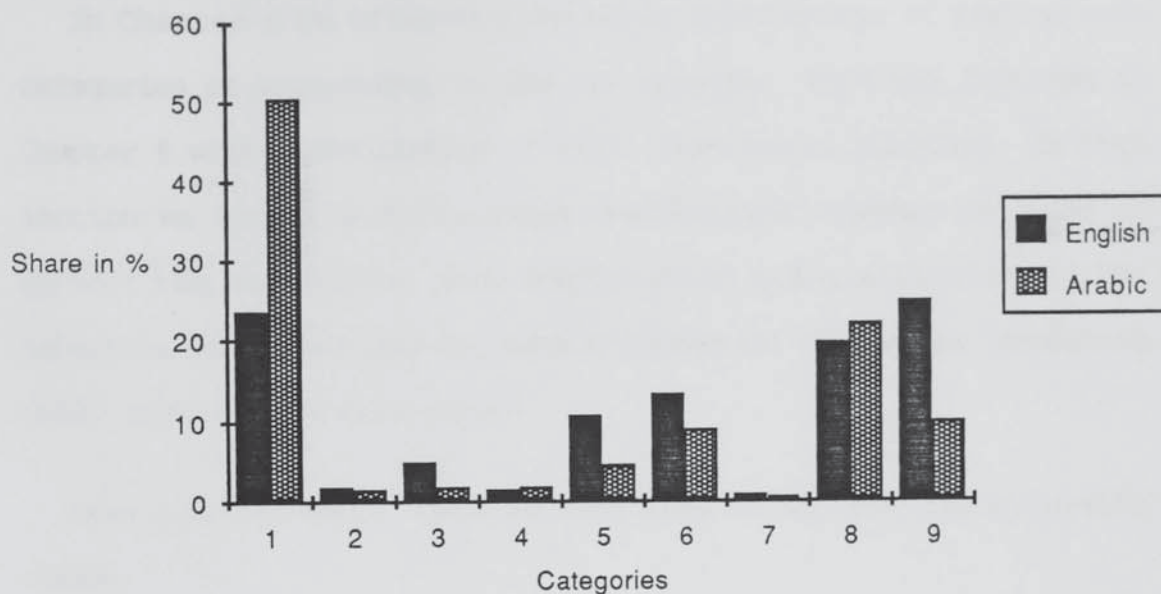
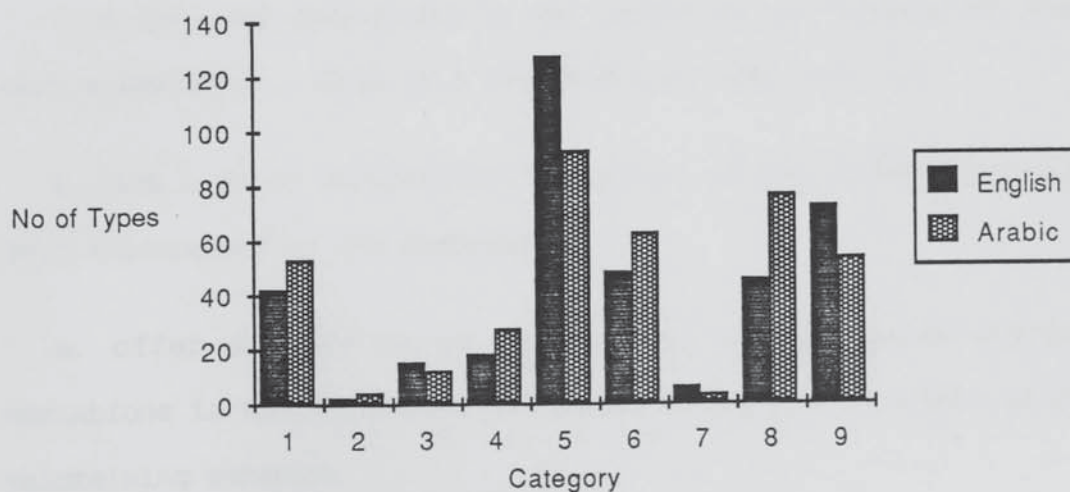


Fig. 9.11 Comparison of the Distribution of Connective Types within Categories of Connectives



1. Additive 2. Alternative 3. Comparative 4. Reformulatory
5. Orientative 6. Temporal 7. Spatial 8. Causal 9. Adversative

9.2 Interlingual Variations in the Functional Categories

9.2.1 Introduction

In Chapter 6 we offered a detailed description of the various categories of connectives in the two corpora. This was followed in Chapter 8 with a description of their statistical profiles. In this section we intend to bring these descriptions together in order to exhibit the variations, both quantitative and qualitative, in the behaviour of connectives and make differential statements concerning their role in each text corpus.

More specifically, this section aims to achieve the following tasks:

1. A general statistical comparison for each category of connectives across the two corpora. The aim is to outline the main differences in frequency, repetitiveness and growth.
2. A detailed comparison of the behaviour of connectives within each subcategory. This is a two-fold task that aims to:
 - a. give a brief statistical comparison of the connectives within each subcategory in the corpora;
 - b. offer an account of the textual (functional/rhetorical) variations in the behaviour of connectives and in their role in maintaining cohesion.

For the sake of consistency, the order of presentation of each categorial comparison will follow the one offered in Chapter 6. Each contrastive account will start with outlining the statistical variations to offer a background for the textual variations

discussed afterwards.

9.2.2 Additive Connectives

9.2.2.1 Quantitative Variations of Additive Connectives

The most striking difference in the quantitative distribution of the various categories of connectives in the two languages lies within the category of additive connectives. Every measure that we have used reveals statistical variations across the two corpora. These are discussed below:

1. Frequency Distribution

The first striking difference is the frequency distribution of additive connectives in English and Arabic. The following differential statements give a summary of the variations:

a. The total number of additive connectives in Arabic is 8,552 tokens, which represents 50.32% of total token coverage. In comparison, additive tokens in the English corpus are 2,260, which represents 23.55% of total tokens. This result exhibits the importance of additivity as a factor in organising text sequences in Arabic as opposed to English. Figures (9.12-13) display a comparison of the distribution of additive connectives within text tokens and within connective tokens in the corpora.

b. Additive connectives in Arabic comprises a bigger number of types (53 representing 14% of types at this level of categorisation) than their counterparts in English (42 types representing 11% of types).

Fig. 9.12 Comparison of the Distribution of Additive Connectives

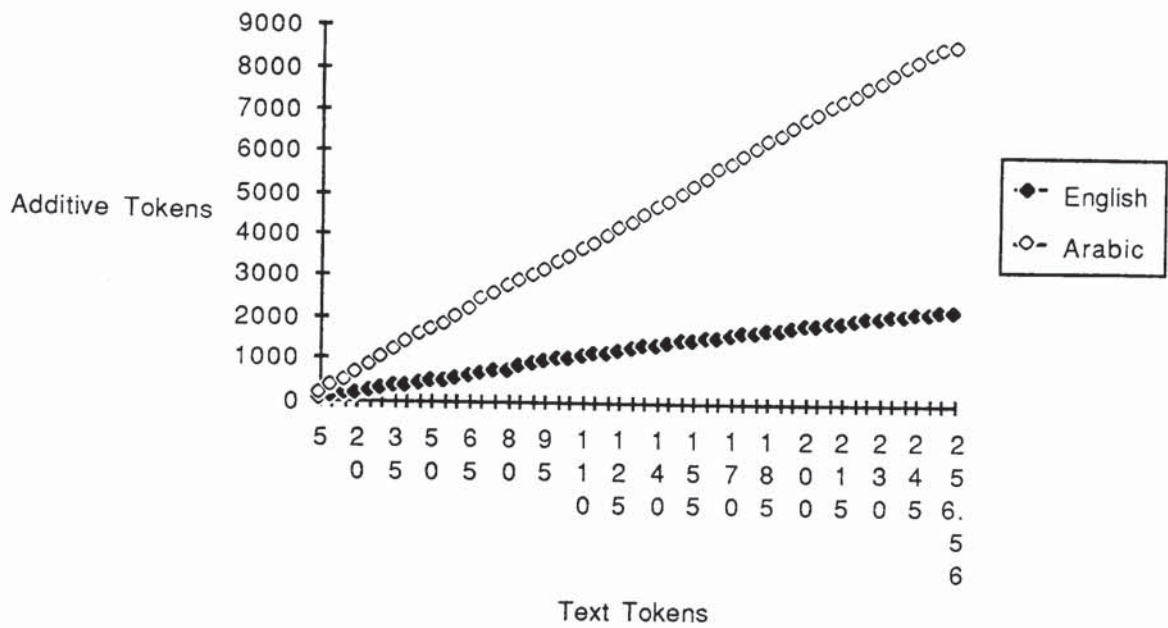
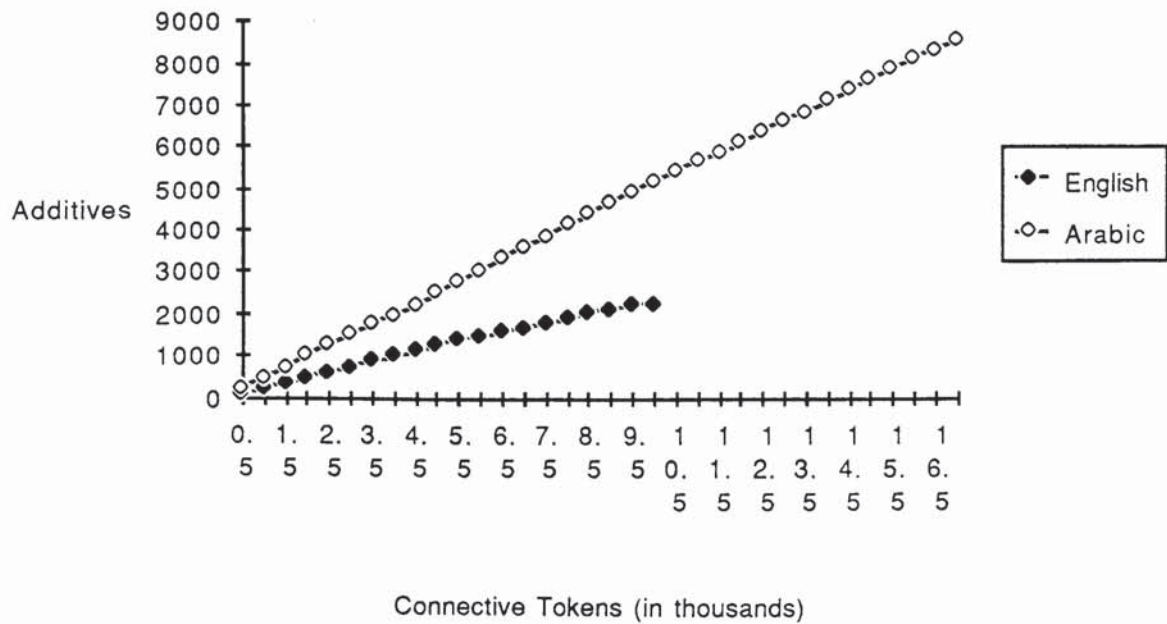


Fig. 9.13 Comparison of the Distribution of Additive Connectives within Connective Tokens



c. Arabic additivity relies heavily on the use of the connectives "wa" and "fa", the total frequency of both representing 89% of additives. This size is represented in English by the combined frequencies of 7 connectives: and, also, even, again, which, too and nor.

2. Repetitiveness

A comparison of the indices of repetitiveness of additives in both corpora shows that the Arabic additives are more repetitive than English. The following statements give a general contrastive summary:

a. Additive occurrence rate is very high in Arabic. Every other connective token in the corpus is an additive. In English an additive occurs in every 4 successive connectives.

b. Type occurrence rate within the category is lower in Arabic than in English. In Arabic a new type appears in every 161 successive additives; in English it appears within 54. This difference is due to the enormity of the number of additive tokens in Arabic compared to English.

c. Both general and system repeat rates are higher in Arabic than in English. The indices show that the probability that two successive connectives turn out to be any additives is 0.25 in Arabic and 0.055 in English. The probability that two successive connectives are the same additive type is 0.172 in Arabic and 0.021 in English.

d. The gap distribution shows shorter distances separating two additives in Arabic than in English. The average distance in Arabic is 29 and in English 112.

3. Growth

Comparison of global and local growth of additives are displayed in Figures (9.14-15).

a. Global Growth:

i. Due to the enormous number of additives in Arabic growth of tokens is larger in Arabic than in English. Examined within intervals of 5,000 words each, maximum growth level is 205 additive tokens in Arabic and 64 in English. The average growth at any interval is 167 additives in Arabic and 44 in English.

ii. Comparison of growth of additive types in the same intervals shows faster growth and earlier saturation in Arabic compared to English. The 50th percentile type occurs within interval 3 (15,000 words) in Arabic and within interval 6 (30,000 words) in English. The 75th type percentile occurs within interval 14 (70,000 words) in Arabic and within interval 18 (90,000 words) in English. The 90th percentile type is encountered within interval 22 (110,000 words) in Arabic and within interval 35 (175,000 words) in English.

iii. Extrapolation of a million word Arabic corpus is expected to include 32,641 additive tokens and types. An English corpus of the same size is expected to contain 9,309 additive tokens and 74 types.

Fig. 9.14 Comparison of the Global Growth of Additive Connectives

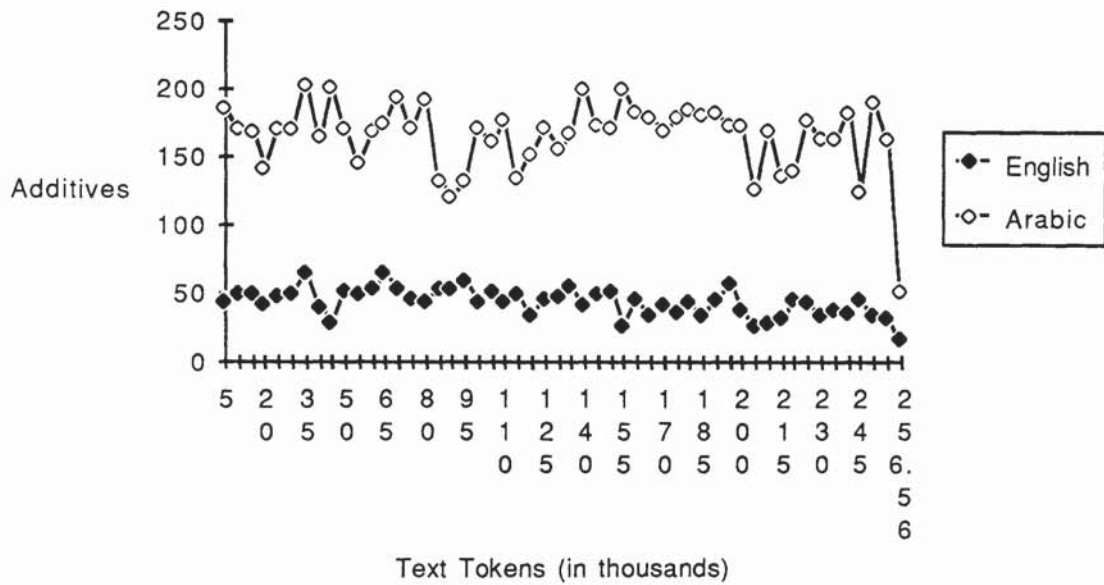
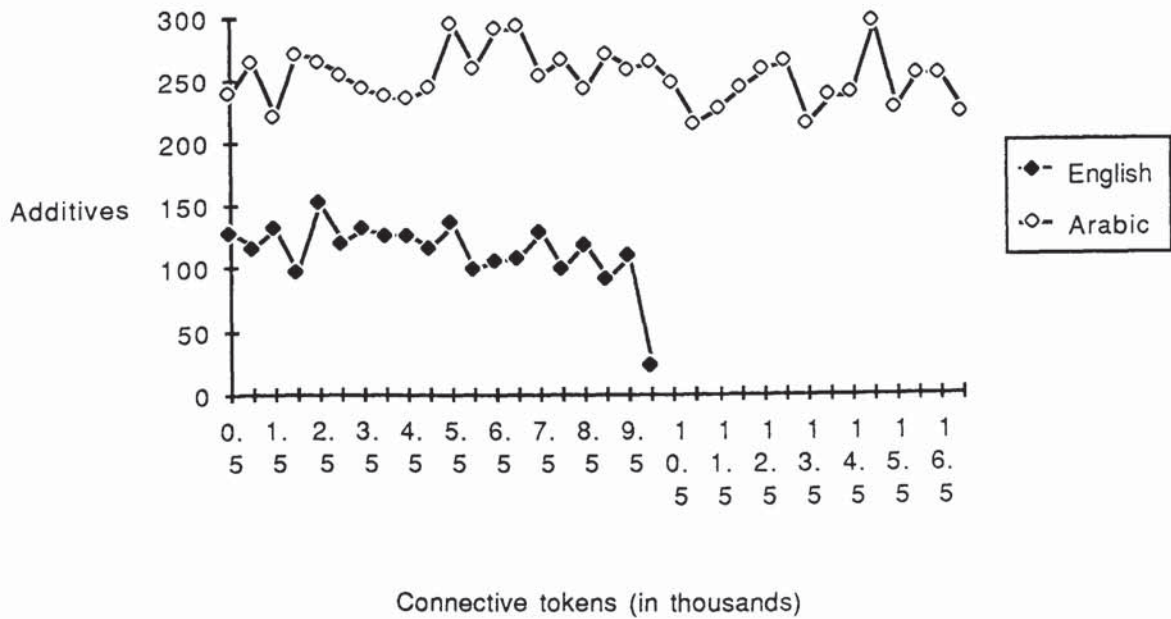


Fig. 9.15 Comparison of the Local Growth of Additive Connectives



b. Local Growth

Comparison of the growth of additive connectives within text intervals each containing 500 connective tokens shows larger growth in Arabic than in English. The maximum growth level is 293 additives (58.6% of connectives in the interval) in Arabic and 152 additives (30.4% of connectives in the interval) in English. The minimum growth level is 214 additives in Arabic and 97 in English. The average growth within any interval is 252 additives in Arabic and 118 in English. Extrapolation of an Arabic corpus containing 100,000 connective tokens (regardless of the size of text in terms of words) is expected to contain 52,012 additives. A similar size English corpus is expected to contain 23,267.

4. Distribution of Additive Categories

Comparison of the distribution of additive categories in English and Arabic is displayed in Figures (9.16-17). A summary is given in the following statements:

a. Appending

This is the biggest additive category in each corpus. However, their size differs in both corpora. In English, it comprises 1,805 tokens, a smaller set than its Arabic counterpart, which consists of 4,058. However, the share that this category represents in additive mass is greater in English than Arabic: 80% compared to 47%. This points to the reliance of additive connectives in English on appending and the comparative small size of other categories of English additives.

Fig. 9.16 Comparison of the Distribution of Categories of Additive Connectives in the English and Arabic Corpora

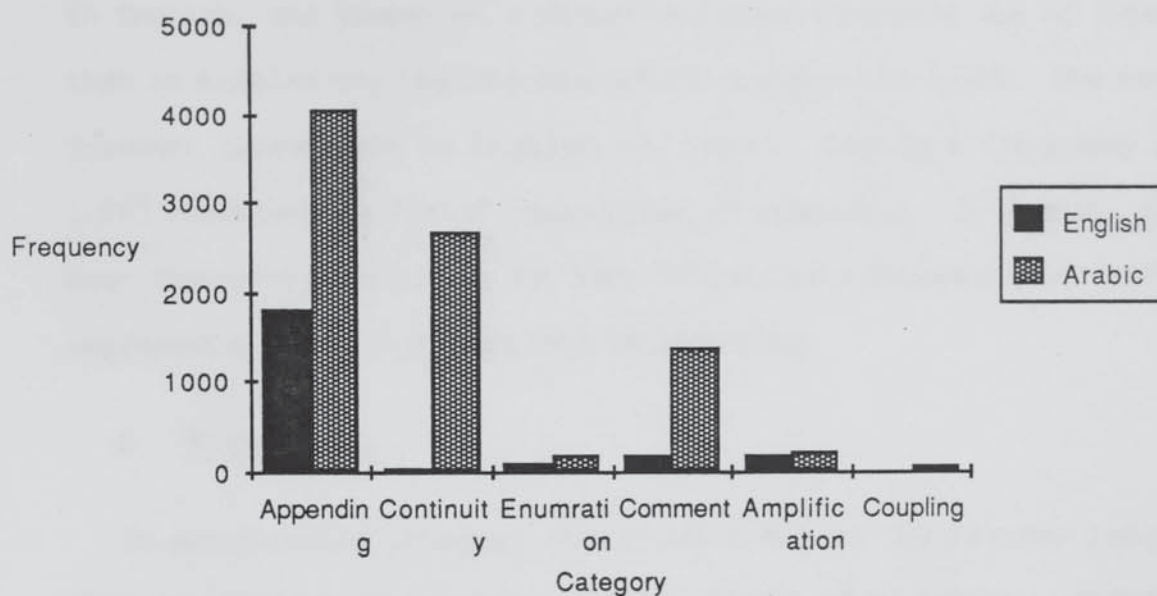
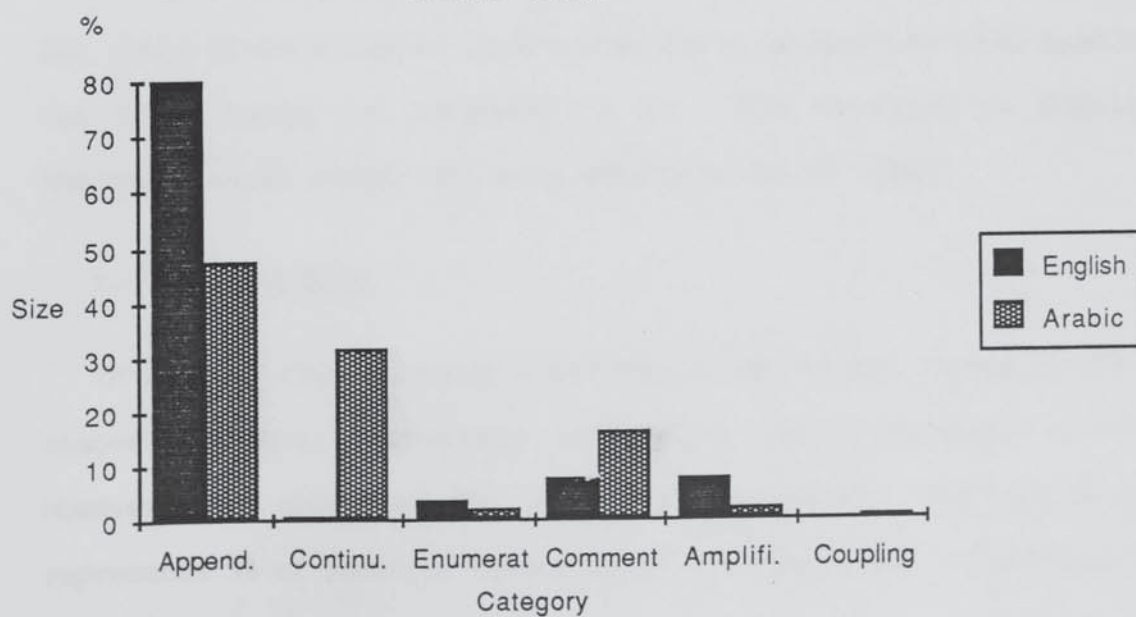


Fig. 9.17 Comparison of the Distribution of Shares of Additive Categories



The number of types at this level of categorisation is similar in both corpora, both having 21. This means a higher type-token ratio in English, and therefore a richer and more extensive use of types, than in Arabic; the indices being 0.01 compared to 0.005. The most frequent connective in English is "and", having a frequency of 1,263 representing 70% of connectives of appending. In Arabic, the most frequent connective is "wa", which has a frequency of 3,479, representing 86% of connectives of appending.

b. Enumeration

In Arabic this category comprises a set of 180 tokens, larger than its English counterpart, which consists of 88 tokens. However, the share that each has among additives is bigger in English; it is equivalent to 4% of additives while in Arabic the share represents 2% of additives.

The category in English comprises 18 types; in Arabic it contains 20. This gives a higher type-token ratio in English than Arabic, the index being 0.2 compared to 0.1. The category in English therefore shows richer and more extensive use of types.

c. Amplification

In English this category comprises a set of 204 tokens slightly bigger than its counterpart in Arabic, which consists of 178. However their share in additive connectives differ. The English set represents 9% of additive tokens while the Arabic set corresponds to 2%.

The number of types in each set is nearly similar, 7 in English and 8 in Arabic. This gives type-token ratios of 0.034 in English and 0.045 in Arabic. The most frequent connective in English is "even" which has a frequency of 142, equal to 80% of tokens in the category. In Arabic 3 connectives achieve this share in token coverage: "bal" (119 occurrences, 58% of tokens), "ḥattā" (36 occurrences, 18% of tokens) and "innamā" (18 occurrences, 9% of tokens).

d. Comment

Additive connectives of comment constitute a significantly larger set of tokens in Arabic than in English. It comprises 1,380 tokens within the Arabic but only 173 in English. This is indicative of the need in Arabic text for explicit signalling of operations of regression where the content of the current text sequence is evaluated, defined or disambiguated. The Arabic set represents 16% of additives while the English set achieves less than 8% of additive coverage.

Each set comprises a small number of types. In English there are 3 types: "and", "which" and "after all". In Arabic, the types are 5: "wa", "fa", which both achieve 99% of token coverage, "alā wa", "mimmā", and "ḥaytu".

e. Continuity

This a larger additive category in Arabic than in English. The Arabic corpus contains 2,627 tokens representing 31% of additives. It comprises 7 types, the most frequent is "wa", which stands for

99% of continuative tokens. In contrast, only 16 continuatives have been identified among additives in the English corpus, a frequency that corresponds to only 0.7% of additives.

f. Coupling

This set has been identified in Arabic only. It comprises 54 tokens and 3 types: "fa" (24 occurrences), "wa" (21) and "'aw" (9).

9.2.2.2 Textual Variations of Additive Connectives

Additive connectives do not only exhibit sharp quantitative differences in the two corpora, but also some distinct textual variations in the way the various functions that are here grouped as additives are utilised. To explore and catalogue these differences, we shall look at each additive subcategory and examine where textual tendencies seem to diverge in the corpora.

9.2.2.2.1 Appending

Variation in the use of the textual patterns of appending can be summarised in three: flexibility, the strong-weak additive distinction and the use of parallelism.

1. Flexibility

An interesting point of departure in the patterns of appending in the two corpora is the flexibility of satisfying the condition of sufficiency for their use in Arabic as compared to English. A sufficient condition for appending to operate in Arabic is the existence of two propositions, or blocks of knowledge, that have some form of ordering in the textual world, no matter how fuzzy,

approximate or inconsistent this form may be. Even the slightest indication of ordering is sufficient. Once the condition of sufficiency is satisfied, it immediately becomes necessary to annex the two propositions, whether in the absence or even presence of other connectives, via an appending connective, the most frequent being "wa". This factor accounts for the enormity of the number of appending connectives used in Arabic.

Rhetorically, the absence of appending in this case causes a "gap" between the two text sequences that is damaging to the efficiency of text cohesion. In terms of processing, this gap causes the processor to halt immediately and inspect the options available for establishing a link. Normally the processor will track back and re-process the current block of knowledge. If the expected link is still that of appending and the connective is missing the processor will either get "worried", or track back looking for more (possible rhetorical) clues. This will lengthen processing unnecessarily.

In comparison, the extent of flexibility in satisfying conditions of sufficiency that is available in Arabic is non-existent for English (with reference, of course, to the corpus). First, ordering the facts or blocks of knowledge has to be clear in the text world, and, secondly, an incremental effect is to be expected. But even when these two conditions of sufficiency are satisfied, it is, unlike Arabic, not necessary to annex the two blocks of knowledge, and the option is left open for text utilisation:

- a) The relation can be left implicit, particularly across the

sentence or larger sequences of text. This implicitness, as effected through the absence of an appending connective, is compatible with the organisation of blocks of knowledge. The processor, after advancing through, and successfully completing processing of one block, will, in the absence of a connective that is indicative of the relation between the current and subsequent blocks, assume an appending relation and start to break and span the next block on this basis. If the content of the next block confirms this relation, the processor is enabled to move on. If there is no confirmation the processor will track back to initiate the correct relation and selects the appropriate path.

b) The relation can be signalled explicitly. Explicitness is necessitated when i) ambiguity may arise that can damage the implicit expression of appending, or, ii) when incremental and/or ordering effects are intentionally created for a more forceful expression of additivity.

2. Strong-weak connective distinction

Appending connectives in Arabic are divided into two distinct types according to their force of connectivity.

a. Strong: This type include connectives that have sufficient force of connectivity for performing appending. These are usually positioned initially in relation to the subsequent clause they append. The group includes "wa", "fa", "tūmma" and "kamā".

b. Weak: This subsumes a group of connectives of two kinds:

i) The weaker group cannot perform appending on their own and, therefore, require the additional use of a strong connective, particularly "wa". To this group belong the connective "'ayḍan" and "fawqa (ḍālika)". the position of "ayḍan" is medial or final within the subsequent clause; the position of "fawqa ḍālika" is initial, but has to be preceded by "wa".

ii) The majority of weak appending connectives, however, may occur on their own, but they are rhetorically felt to lack sufficient force to create efficient connectivity. Hence they are often preceded by a strong connective, particularly "wa". To this group belong such connectives as "kaḍālika", and "marratan 'uxrā". Their position is normally initial in the subsequent clause, often preceded by "wa".

This strong-weak distinction of appending connectives is blurred and not very clear in the English corpus.

3) Parallelism

One of the recurrent functions of appending in Arabic is to annex, and therefore support the use of, parallelistic forms. This point will be discussed with some more detail later. It suffices here to state that this function of appending seems more distinct in the Arabic than English corpus. The following example shows the extent to which parallelistic forms can be used to support an argument.

[9.1] wa ḥina' iḍin sa-yaruddu al-ulamā'u
 wa yaṭūru al-'awliyā'u
 wa yatawassaṭu al-ru''asā'u
 wa yamurru al-'ahammu alladī yurīd-ūna
 wa al-nāsu maṣḡulatun bi-al-gammi alladī yuṭliq-ūna.

[... then scholars of theology will argue to disprove, the clergymen will be infuriated, and the rulers will interfere; and then these will pass whatever laws or decrees they like, while people are engaged in gloomy argumentation ...]
(Nb, 4/4/83, X3, 239-243)

Note that the previous text sequence (all forming one sentence) is composed of 5 clauses that form two groups: one comprises the first three propositions and the other the last two. Within each group, the clauses are in parallel to each other. The structural parallelism within each group is intensified by the use of identical rhyme and similar rhythm. The connective "wa" controls the ordering of these forms within and across the two groups.

9.2.2.2.2 Enumeration

The exposition of the textual functioning and patterning of enumeratives in the two corpora is discussed with extensive exemplification in 6.4.4 above. Included is also a basic contrastive description of the behaviour of connectives in the two languages. However, here we shall concentrate on describing similarities and variations with some more detail.

Rhetorically, additive connectives of enumeration in both languages organise text in distinct sequences that intend to reflect a division on the basis of listing of comparable items (points, arguments, etc). The connectives act as signposts that identify and limit the sequence and suggest an ordering plan or reveal its details more closely. Signposting, it should be noted, is to be distinguished from topical development. The latter makes an assertion around the subject or topic; signposts, on the other hand,

manifests the organisational framework of the discussion of the topic.

In terms of processing, enumerative connectives indicate a type of recursion where one category in the linear sequence of text is selected and repeated. The processor responds by reactivating its operations for each enumerated item until the last terminates. To adapt to recursion, the processor employs elaborate looping operations of two types: of regression and progression, (see Chapter 4) which has the advantage of cutting on load.

This textual tactic, which is initiated and supported by the appropriate connectives, differs in both corpora, both in terms of rhetorical effect type of processing operation involved. The differences reflect tendencies rather than absolute variations in the way text sequences are organised and integrated.

The main observed difference concerns "narrow" vs. "broad" ordering of enumerated sequences. The tendency in Arabic, and this is evident in the corpus, is towards "narrow" ordering, where a tight plan is created through systematically numbering the sequences. Connectives that support this plan include "awwalan" [first], "ṭaniyan" [second], "ṭaliṭan" [third], and so on. Numbering can go on and there is a case, as mentioned earlier (Chapter 6), where it reaches "ḥādī <aṣar" [eleventh]. This ordering facilitates accessibility of text organisation and creates a tightly discrete itemisation that connectives consistently signal until the last item in the order is utilised.

In contrast, there is a strong tendency in English for a "broad"

type of ordering, where the components of a plan is often signposted by such items as "first", "then" and "lastly". Although these connectives still serve in text utilisation as framing devices that remind us of the framework of the plan, they are more flexible in the way the various items are grouped or introduced.

Within the perspective of processing, the variation in the tendencies represents a difference in tactic. In "narrow" ordering, there is a clear domination of the operations of progression, since the cataloguing of the items is evident at once. The connectives help the processor to adapt to these operations, for instance by anticipating a loop and successfully advancing forward. On the other hand, "broad" ordering requires operations dominated by regression (though, of course, progression is still present). Although the connectives help the processor to manipulate progression strategies, since they label the loops and help advance it further, the processor has to extract the shared criteria of the list after consulting the previous items or, in some cases, most of the list items.

9.2.2.2.3 Amplification

One main interlingual difference in the textual functioning of connectives of amplification lies in the textual patterns of the two most frequent connectives: "even" as opposed to "bal". Their cohesive role can be assessed by examining two aspects of their textual function: ordering and intensification.

Both connectives organise propositions in a particular order that

does not necessarily correlate with the real order of events and situations. The conceptual configurations that underlie the surface expression of propositions are graded according to a scale. Starting from the lower end of the scale propositions are graded additively. However, to override the smoothness of the ordering, the connectives "even" and "bal" signal a move to a higher grade on the scale. It is in the way this move is achieved that the two connectives differ. In Arabic, "bal" often marks a sharper, more abrupt and higher move towards the end of the scale and therefore achieves a more acute degree of unexpectedness. This move can be intensified further by the use of "wa" in an additive function of comment. Thus, while the two connectives support similar linearised ordering, they depart in the degree of intensifying the content of the subsequent propositions.

In this way, the connective "bal" assists in reflecting a bigger share of interaction between text utilisation and other human sensory modes. To elaborate this statement, we should note that humans distribute attention selectively in order to achieve a good retention of knowledge configurations. The sharp degree of amplification that "bal" signals makes it a concrete cue of presentation that forcefully intrude on sensory apperception, and thus best noticed and recalled. This is an essential factor in achieving textual salience and "bal" seems to express it far more effectively than "even".

9.2.2.2.4 Comment

Additive connectives of comment, as described in 6.6, carry out

two distinctive functions: the first is "comment proper" and is shared by the two languages; the second involves creating focus and is peculiar to Arabic. This immediately establishes the functional convergence and divergence in the use of this category of additive connectives in the two languages. This is discussed with more detail below:

1. Comment "Proper"

Rhetorically, connectives of this category create a pause and allow the text producer to examine what he has said and offer a comment, definition, or clarification of all, or just one expression within, his previous statement. From the viewpoint of processing, the connective signals a number of operations dominated by regression whereby the processor is involved in activities of expanding or modifying the conceptual configurations for the previous surface stretch of text.

2. Focus

The main observed interlingual variation within this category lies in the second function of the comment connectives: that of creating focus. This involves two related operations that have an important role in organising text sequences and maintaining cohesion in Arabic text. Both operations are not required in English text.

A. Bridging discontinuities

In Arabic discontinuity occurs whenever a sentence starts with an adverbial phrase, including those that function as connectives, and followed by a nominal clause or a clause introduced by "inna". One

reason that accounts for this discontinuity is that the conceptual configuration of the opening phrase seems to impose a sizeable load upon processing to the extent that the processor is made temporarily inactive. The connective assists in a) activating the processing operation so that a path is immediately taken, and b) drawing control towards the conceptual content of the proposition itself. Examples from the corpus (each is the first part of a new sentence):

- [9.2] wa bi-al-tālī fa 'inna-hā tajidu nafsa-hā
 <ājizatan ...
 [And then she finds herself unable to ...]
 (Ar, 21/1/83, X180, 31499-500)
- [9.3] wa bi-al-ṭab<i fa huwa bi-jānibi al-kaffati
 al-rājiḥati.
 [And, of course, he is with the winning side].
 (Ar, 21/1/83, X179, 31395-6)
- [9.4] 'idan fa <alay-nā 'an nuṣalliḥa hādā al-xaṭa'
 [Therefore we have to rectify this mistake ...]
 (Ar, 13/6/83, X205, 35431-2)

A similar discontinuity occurs when, instead of an adverbial phrase, the sentence starts with an adverbial subordinate clause. The connective "fa" is required to bridge a gap that separates the main from the subordinate clauses and put the main clause into focus. The following are some examples from the corpus:

- [9.5] <indamā yajhalu al-'insānu šay'an fa 'inna-hu
 yaxšā-hu.
 [When man is not familiar with something, he
 dreads it]
 (Hr, 25/1/83, X33, 6462-3)
- [9.6] wa 'ayyan kāna al-'ixtiyāru, fa 'inna hādihi al-
 jamā<ata taqūmu al-yawma bi-kitābatī tārīxin
 jadīdin la-hā ...
 [Whatever the alternative, this faction (of the

Palestinian resistance) is starting a new history
for itself ...]

(Nb, 7/7/83, X17, 3252-3)

In English, this type of text linearisation does not create discontinuity. Hence in text production there is no requirement in such cases to resort to explicit connective use for the purpose of filling a gap and relating sequences while creating focus.

B. Intensification

A related function that is discussed in 6.4.6 and involves mainly the connective "wa" is that of "intensification". Here the connective not only brings the next proposition under focus, but acts to intensify the semantic scope of an immediately preceding expression, usually one of a certain number of connectives. Examples from the corpus:

la budda wa 'an ...
'ilā hunā wa ...
xāssatan wa ...

This way of linearising sequences for achieving a degree of salience is not present for English.

9.2.2.2.5 Continuity

The differences in the textual functioning and patterning of additive connectives of continuity in the two corpora have already been discussed (see 6.4.7 above). We shall, therefore, stress here the distinctive variations in the way these connectives sustain text cohesion. These variations are discussed under two points: explicit signalling and connective support.

1. Explicit signalling

Additive connectives of continuity mark a sharp distinction in text cohesion in both English and Arabic. In English, these connectives play a small and restricted role. Text sequences of all sizes in English assume continuity without the requirement of explicit signalling. In other words, additive continuity is not necessarily explicit, nor are text sequences necessarily related, in the absence of other types of signalling, by additive connectives of continuity. Indeed, if continuity is imposed upon text sequences by consistently annexing them via additivity, text stability will be adversely affected, as the entire procedure will run contrary to normal text organisation.

One reason for this is that in text organisation in English, unbroken access among text sequences is maintained via various other signals of cohesion. The number and type of signals that are needed for continuity to prevail are determined by such factors as intended rhetorical effect, pressure of text organisation and the special requirements of context.

In Arabic, however, continuity via additive connectives is a major means for organising conceptual configurations. For despite the occurrence of various other signals of cohesion, text sequences - particularly, the sentence and larger sequences - remain discontinuous unless the subsystem of additive connectives of continuity is utilised in the actualisation process.

It should be noted that where a topic shift is intended, whether within or across the paragraph boundaries, the option is left open

to utilise or disregard additive connectives. The absence of a connective is tolerated on the basis that it designates conspicuously that a new topic, a new set of knowledge configurations in the text world, is to be initiated. However, this can often be overridden, and the Arabic corpus contains numerous examples where this marking of topic shift is sacrificed for achieving tighter continuity.

2. Connective Support

It is often the case in the English corpus that connectives of various categories are characterised by self-sufficiency in relating and organising various conceptual configurations. This characteristic is both inherent and acquired within connectives: a) It is inherent by virtue of the semantic space which makes it capable of integrating two propositions (or sets of propositions) into a larger sequence for use in further processing and for both active and long-term storage. b) It is acquired because other types of cohesive devices and the way text is linearised in the process of text actualisation contribute to the efficiency of the connective in integrating text sequences.

This property gives connectives a cohesive force of two overlapping dimensions: i) they can relate text sequences by signalling the type of semantic relation that combines one to the other, and ii) they can integrate sequences by imposing continuity. Hence connectives in English do not require support to signal continuity.

The situation is not the same in Arabic. Cohesive self-sufficiency for most connectives, excluding a small group that include "wa", "fa", and "tamma", is circumscribed by two related factors:

a) The semantic space is capable of encompassing two propositions and denominating the type of relation that supervenes. However, it cannot secure continuity within a larger 'chunk' of text.

b) The connective, therefore, has to rely on the support of its environment for the acquisition of textual continuity. This is normally provided by complementing the textual role of the connective with an additive connective of continuity.

To verify this statement we used OCP to generate all occurrences of additive connectives of continuity that support the cohesive function of other connectives in the two corpora. The results prove that a large percentage of connectives may require additive support, depending on a number of textual factors. These will be mentioned where appropriate in this contrastive exposition.

9.2.2.2.6 Coupling

Coupling is a type of conceptual isomorphism. Additive connectives are occasionally used to connect two propositions with identical or similar conceptual content. The fact that this relation is predominantly peculiar to Arabic reveals a further aspect of text organisation in Arabic as opposed to English, a point that will be expanded later (see Chapter 10).

In general, additive connectives of coupling activate operations

of regression whereby a conceptual configuration is reiterated. This may cause some loss in terms of processing time (as the system slows down) through reactivating the same previous semantic network. However, the loss is immediately counterbalanced by gain through less intensive processing, since the input is practically unchanged, and through consequent generation of greater concentration for the conceptual content. This helps achieve salience, which is the original motivation for coupling.

9.2.3 Comparative Connectives

9.2.3.1 Quantitative Variations of Comparative Connectives

Comparative connectives in the two corpora manifest variations in their distribution, repetitiveness and growth. These are discussed in turn.

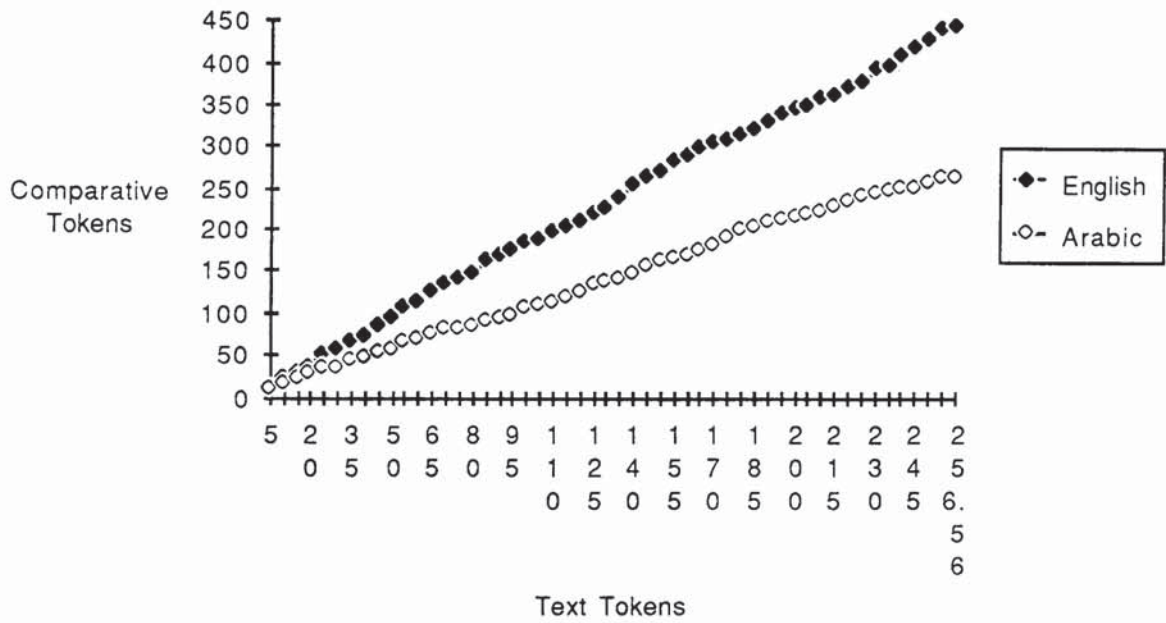
1. Frequency Distribution

The following differential statements summarise the main variations in frequency:

a. The total number of comparative connectives in English is 441 tokens, which represents 4.6% of connective token coverage. In contrast, Arabic has 266 tokens that correspond to 1.6% of token coverage. Figure (9.18) shows a comparison of the distribution of connectives within the text corpus. Figure (9.19) plots the distribution of comparative connectives against connective tokens.

b. The English sets of comparative connectives comprises 15 types while its Arabic counterpart contains 12 types. The type-token ratio is 0.034 in English and 0.045 in Arabic. These figures imply

Fig. 9.18 Comparison of the Distribution of Comparative Connectives



relatively richer use of types in Arabic compared to English.

c. The comparative relation relies heavily in English on the use of the connectives "as" and "than", the combined frequency of which constitutes 78% of total comparative connectives. In contrast, the category relies in Arabic on "kamā" [as] and "ka'nna" [as if], which, together, make up 73% of comparative tokens.

2. Repetitiveness

The following statements give a brief comparison of repetitiveness in both corpora:

a. Comparative occurrence rate is 22 in English and 64 in Arabic. This shows higher occurrence rate in English: one comparative is expected in 22 consecutive connectives.

b. Type occurrence rate is 29 in English and 22 in Arabic. This shows a higher rate in Arabic where a new type is expected to appear within 22 comparative tokens.

c. Both general and system repeat rates are higher in English than in Arabic. The indices show that the probability that two successive connectives in the corpus turn out to be comparatives is (211×10^{-5}) in English and (24×10^{-5}) in Arabic. The probability that two successive connectives are the same comparative type is (75×10^{-5}) in English and (24×10^{-5}) in Arabic.

d. The gap distribution indicates shorter distances separating two sites occupied by comparative connectives in English than in Arabic. The average distance is 579 in English and 959 in Arabic.

3. Growth

Comparison of global and local growth of comparative connectives are displayed in Figures (9.20-21). The following are the main variations:

a. Global Growth

i. The bigger number of comparative connectives in English than Arabic is reflected in their growth. Maximum growth level is 16 comparative tokens in English and 12 in Arabic. Minimum growth level is 4 comparative tokens in English and one in Arabic. The average growth is approximately 9 in each interval in English and 5 in Arabic.

ii. Comparison of growth of types in the same intervals manifests that types grow faster in the early intervals in Arabic than English: the 50th percentile type occurs within the first interval in Arabic and within interval 3 (15,000 words) in English. But the fast growth drops in Arabic to a minimum level. This is clear in the position of the 75th and 90th percentile types in both corpora. The 75th percentile type occurs within interval 14 (70,000 words) in Arabic and within interval 10 (50,000 words) in English. The 90th percentile type occurs within interval 31 (155,000 words) in Arabic and 14 (70,000) in English. Saturation occurs in interval 21 (105,000 words) in English, earlier than in Arabic where it occurs in interval 35 (175,000 words).

iii. Extrapolation of a million word English corpus is expected to include 1,562 comparative tokens and 21 types (i.e. growth of 1,121 tokens and 6 types). An Arabic corpus of the same size is

Fig. 9.20 Comparison of the "Global" Growth of Comparative Connectives

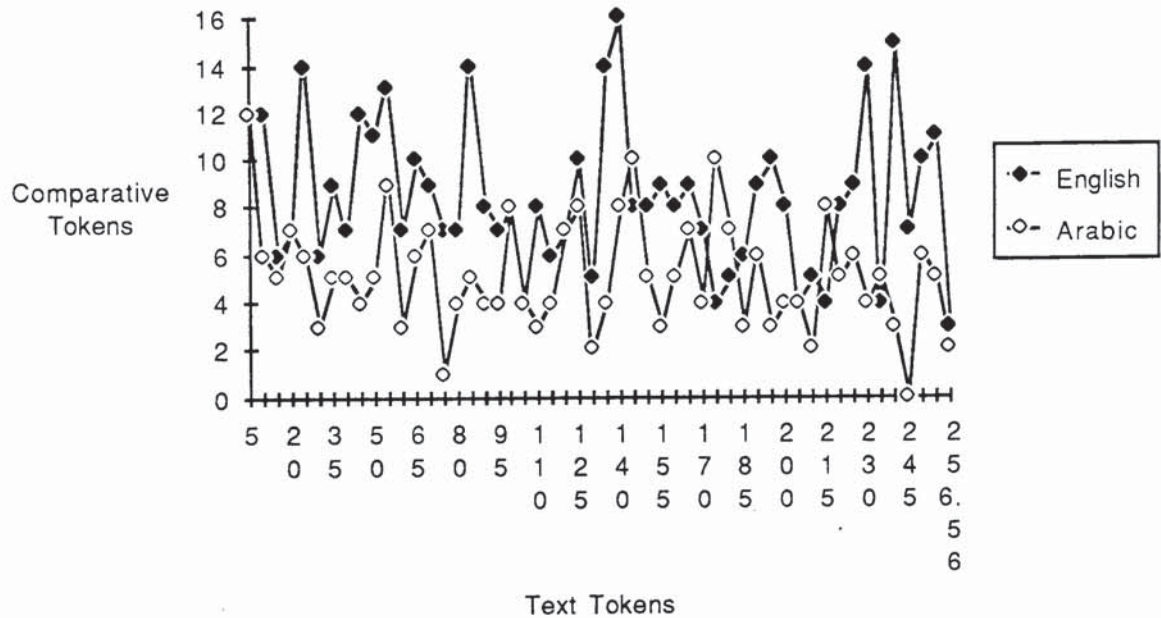
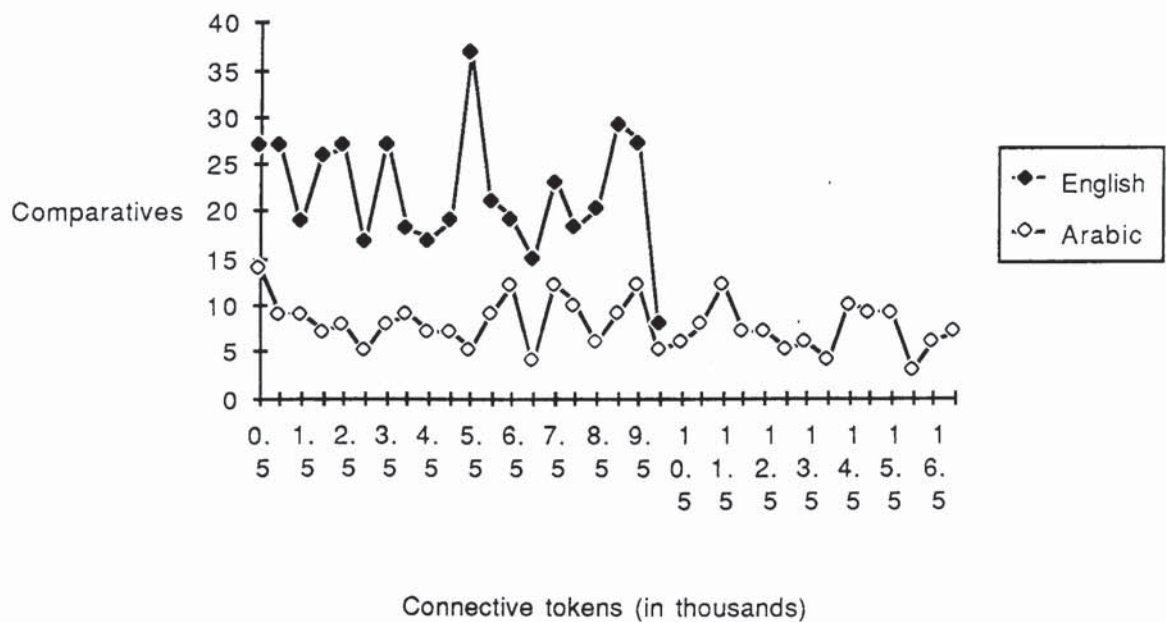


Fig. 9.21 Comparison of the Local Growth of Comparative Connectives



expected to contain 823 comparative tokens and 13 types (i.e. growth of 597 tokens and one type).

b. Local Growth

Comparison of the growth of comparative connectives within text intervals each containing 500 connective tokens shows larger growth in English than in Arabic. The maximum growth level is 37 comparative tokens in English and 14 in Arabic. The minimum growth level is 15 comparative tokens in English and 4 in Arabic. The average growth within any interval is 23 in English and 8 in Arabic.

Extrapolation of an English corpus containing 100,000 connective tokens is expected to contain 3,967 comparative tokens. An Arabic corpus of a similar size is expected to include 1,268.

4. Distribution of Comparative Categories

Comparison of the distribution of comparative categories in English and Arabic is displayed in Figures (9.22-23). The following statements give a brief summary.

a. Similarity

The number of tokens in this category is similar in both corpora: 233 in English and 231 in Arabic. But the share that the category has in comparative connective mass differs. In Arabic it represents 87% of comparatives while in English the share is 53%.

The number of types is also similar, each corpus having 11. But their distribution differs. Arabic relies on 3 connectives "kamā",

Fig. 9.22 Comparison of the Distribution of Categories of Comparative Connectives in the two Corpora

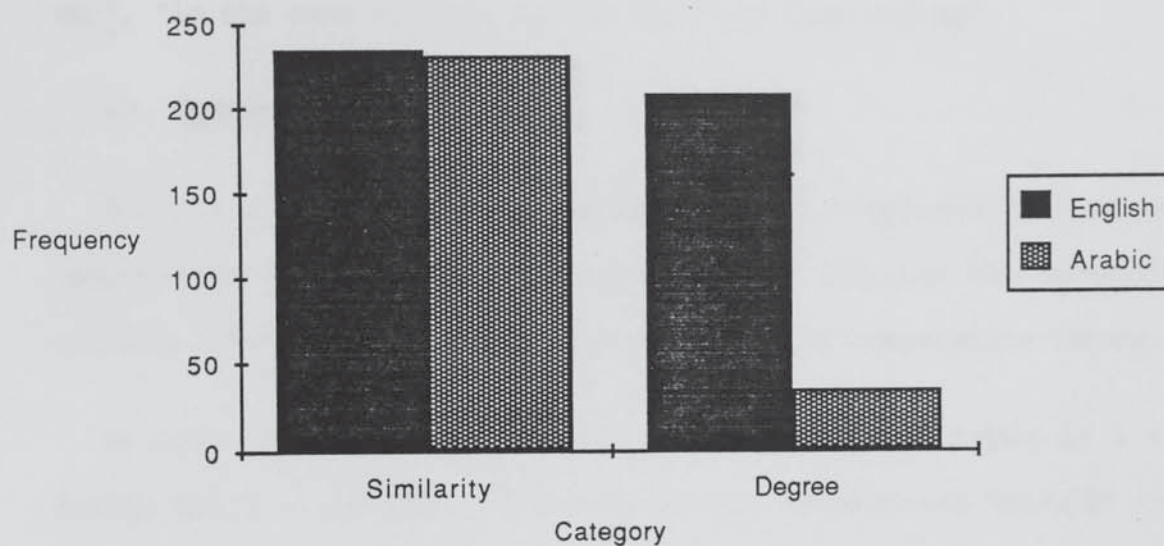
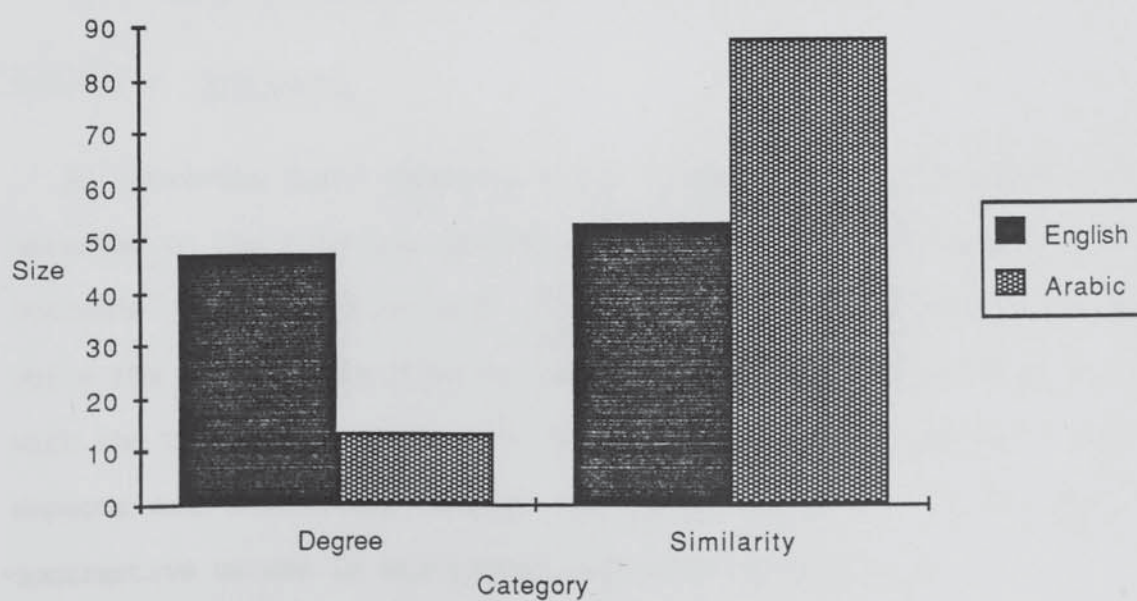


Fig. 9.23 Comparison of the Distribution of Shares of Comparative Categories



"ka'anna" and "mitlamā" which collectively achieve 91% of token coverage. This percentage is achieved by the combined frequency of "as", "as if", "as though", "equally", "similarly" and all connectives where "same" is a core: "the same way", "in the same way", "in the same way as" and "in much the same way as".

b. Degree

This is a small category in Arabic. It comprises 35 tokens, representing 13% of comparative tokens. In English the category consists of 208 tokens, corresponding to 47% of comparative tokens.

At this level of categorisation, the number of types is 4 in Arabic and 9 in English. The two Arabic connectives "mimmā" and "bi-qadri mā" represent 94% of comparative connectives of degree. In English, a similar token coverage is attained by the combined frequencies of "than", "as ... as" and all connectives where "extent" is a core, e.g. "to this extent" and "to that extent".

9.2.3.2 Textual Variation of Comparative Connectives

9.2.3.2.1 Similarity

In achieving their cohesive role, connectives of similarity, as observed in the corpora, differ with respect to two aspects: one concerns linearisation and involves directionality of sequencing, while the other is related to textual functioning and associated with the typology of functions that this category encompasses. Both aspects are mentioned earlier (cf. 6.5) but their interlingual contrastive nature is elucidated with more detail below.

1. Directionality of Functioning

Connectives of similarity in English are of two types. The first evokes operations that are dominated by progression. They comprise the connectives "as", "as if", "as though", "in the same way as", and "in the same way that". These connectives trigger activities of anticipation for the conceptual content within the subsequent propositions. The system load, which can often be heavy in progression, is here eased down since the anticipated content is readily available in the next sequence of text.

The second type signals activities that are dominated by regression. It includes such connectives as "equally", "likewise", "similarly", "by comparison", "in similar fashion", "in a similar way", "in the same way" and "in this manner". They indicate that processing of the next propositions is influenced or controlled by previous activities within the text. Normally the connective elicits intentional reactivation of a previous conceptual network, which then directs the processor to access the subsequent knowledge configuration.

In Arabic, however, the operations that are evoked by connectives of similarity are, generally speaking, mono-directional: they are dominated by progression and therefore the current conceptual content is influenced by the subsequent. However, that should not imply that in Arabic there does not exist any means for signalling operations of similarity that are controlled by regression. Such a means is available, but they are either a) more complex, at least on the surface text, than our definition of connectives will permit, b)

have not been observed to express a prevalent function of similarity, or c) constitute too rare occurrences to establish as a norm. To exemplify the second point, the connective "similarly" have, as a nearest equivalent, the connective "kadalika". But the latter expresses a clear semantic blend of additivity and similarity where the additive meaning is more dominant, and the connective has therefore been classified as an additive.⁽¹⁾

This difference in directionality influences the manner in which text sequences are linearised. Since operations of regression (which the second type of English connectives of similarity are involved in) may look back to a larger "chunk" of text than the one that operations of progression look ahead to, it follows that a) range of connectivity differs in both types of operations, and hence differs across the two languages, and b) grouping is affected across the two languages since connectives of similarity that signal progression seem to group the antecedent and subsequent components more tightly than those signalling regression. This point will be discussed in more detail, though in general terms, i.e. in relation to all categories, when range of connectivity is contrasted across the two languages.

2. Typology of Functions

In discussing the textual functioning of connectives of similarity, five types of sub-function are outlined (cf. 6.5.2.2). Inspection of the functional-quantitative behaviour of connectives in both corpora reveals that while the functions are all operative in both languages, they are better distributed, and hence manifest

more efficient operationality, in English than in Arabic. In contrast, there are two dominant sub-functions of similarity within the Arabic corpus. These are "comparison proper" signalled mainly by "kamā" and "hypothetical comparison" introduced mainly by "ka'anna".

This difference raises the distinction between extensity and intensity of functioning. In English, connectives of similarity manifests richer and more extensive signalling of relations, while in Arabic they display a limited but more intensively applied set of sub-functions.

9.2.3.2.2 Degree

The quantitative contrastive account of this category of comparative connectives underlies their textual variation in the two corpora. There are a number of comments that we may make on the overall textual differences:

1. In English, various relations of degree can be signalled by connectives: equality, inequality and indeterminate degree. The connective in this case reduces processing by directing the various operations (whether dominated by progression, particularly in signalling equality and inequality, or regression as in the expression of indeterminate degree). In contrast, the capability of the Arabic connectives for signalling these relations is limited. First, there is no observed counterpart in the corpus that signals relations of equality of degree in the same way that English connectives do. Secondly, connectives that signify relations of indeterminate degree are very few and have a small number of

occurrences. The main factor behind this is that Arabic employs other more elaborate and structurally different means of expressing these relations than connectives. The use of these means, mostly prepositional phrases and embedding, exerts some pressure on processing and therefore tends to load the system. Rhetorically, the use of the connective is a more direct and therefore more effective means of signalling than complex embedding.

2. An interesting point regarding the use of "mimmā" in the Arabic corpus as opposed to "than" in signifying inequality is the limitation of the Arabic connective in expressing comparison. This is evident not only in the fact that the frequency of "mimmā" is far smaller than "than", but, as inspection of the relevant concordance suggests, in the "point" of comparison that the connective intends to bring out. Apart from 3 instances, the Arabic connective accompanies the word "'akṭar" [more]; in the other three instances the word is "'ab<ad" in the sense of [further, more]. Although there is no structural (syntactic/morphological) reason that prevents the use of other adjectives (i.e. other lexical items) and therefore establish a variety of comparison points, the only one observed in the corpus is that of "size" or "magnitude".

9.2.4 Alternative Connectives

9.2.4.1 Quantitative Variations of Alternative Connectives

This category comprises a small set of connectives in both corpora. Comparison of their distribution, repetitiveness and growth in both corpora leads to the following findings:

1. Frequency Distribution

a. Although the absolute frequency of Arabic alternative tokens is higher than its English counterpart (196 and 143 respectively), its relative frequency is in fact lower. This category represents 1% of total connective tokens in Arabic and 1.5% in English. Figure (9.24) shows a comparison of the distribution of alternative connectives in the corpora. Figure (9.25) plots the distribution of alternative connectives against connective tokens.

b. The Arabic alternative connectives comprise 4 types, while their English counterparts contain only two. However, signalling the relation is achieved mainly by "aw" in Arabic, which stands for 78% of alternative tokens, and "or" in English, the frequency of which is equal to 88% of alternative tokens.

2. Repetitiveness

a. Due to their higher relative frequency, English alternatives have a higher occurrence rate within connective tokens (67 compared to 87 in Arabic). But since they comprise fewer types, their type occurrence rate is lower (71.5 compared to 49).

b. Both the general and system repeat rates are higher in English. The former index is influenced by the higher relative frequency of alternative tokens in English while the latter is affected by the smaller number of types.

c. Due to the bigger number of alternative tokens in Arabic, the average gap separating two Arabic alternative tokens is shorter, the

Fig. 9.24 Comparison of the Distribution of Alternative Connectives

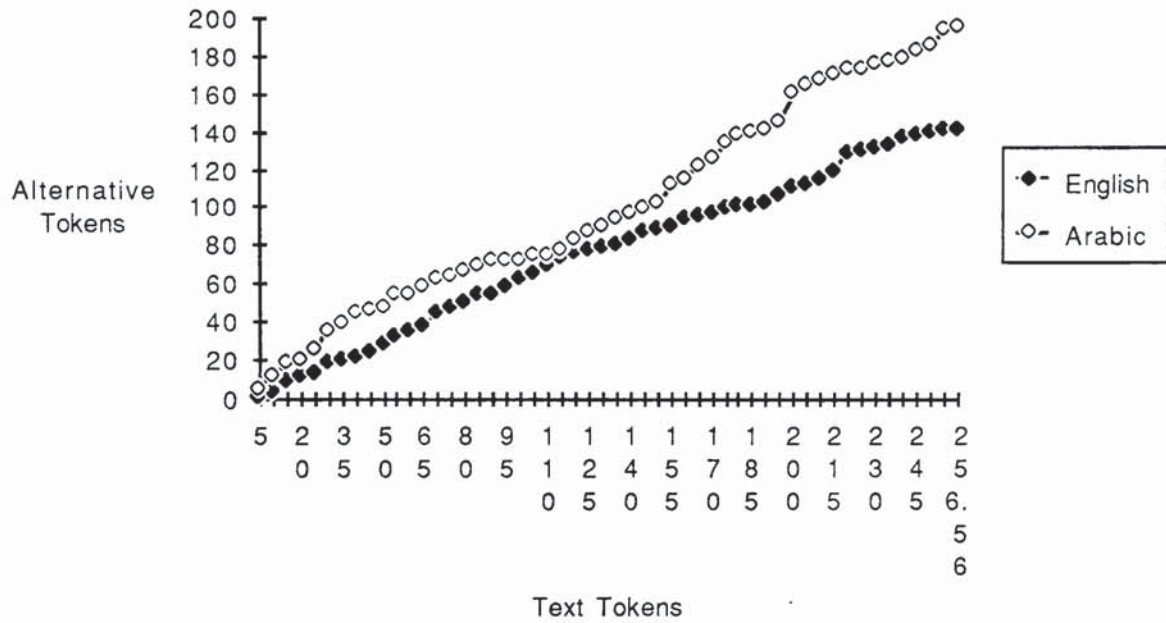
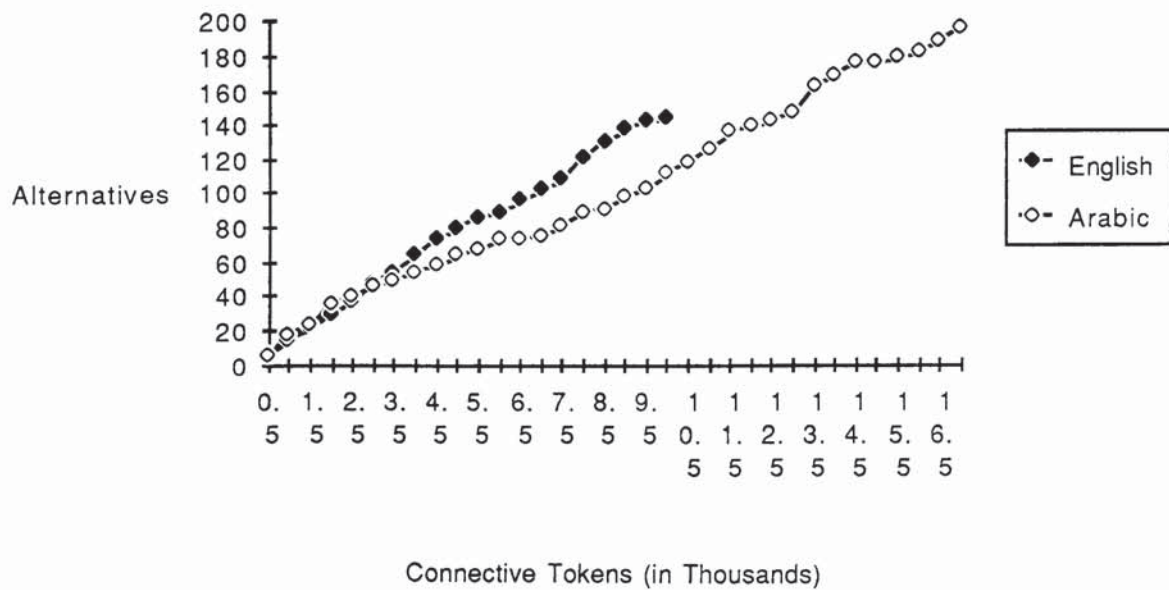


Fig. 9.25 Comparison of the Distribution of Alternative Connectives within Connective Tokens



distance length being 1,300 words compared to 1,780.

3. Growth

Comparison of global and local growth of alternative connectives are displayed in Figures (9.26-27). The following is a brief account of the main variations.

a. Global Growth

i. The average growth of connectives within the intervals is 3 in English and 4 in Arabic. The highest growth level is 9 in English and 14 in Arabic. But despite these higher growth levels in Arabic than in English, there is a bigger growth in tokens in the latter intervals in English than in Arabic. In contrast, Arabic manifests a bigger growth within the first intervals. This fact affects the calculation of growth rate and results of extrapolation. Accordingly, a 500,000 word corpus is expected to contain similar number of alternative tokens, 326 English and 330 in Arabic. A million word corpus is expected to have more alternatives in English than Arabic (respectively 686 and 605).

ii. Growth of types reaches saturation at an early interval in both corpora (interval 4, representing 20,000 words of running text) and no new types appear after that. Accordingly, a corpus of 500,000 or even a million words is not expected to contain any new types in either language.

b. Local Growth

Growth of alternatives within text intervals each containing 500

Fig. 9.26 Comparison of "Global" Growth of Alternative Connectives

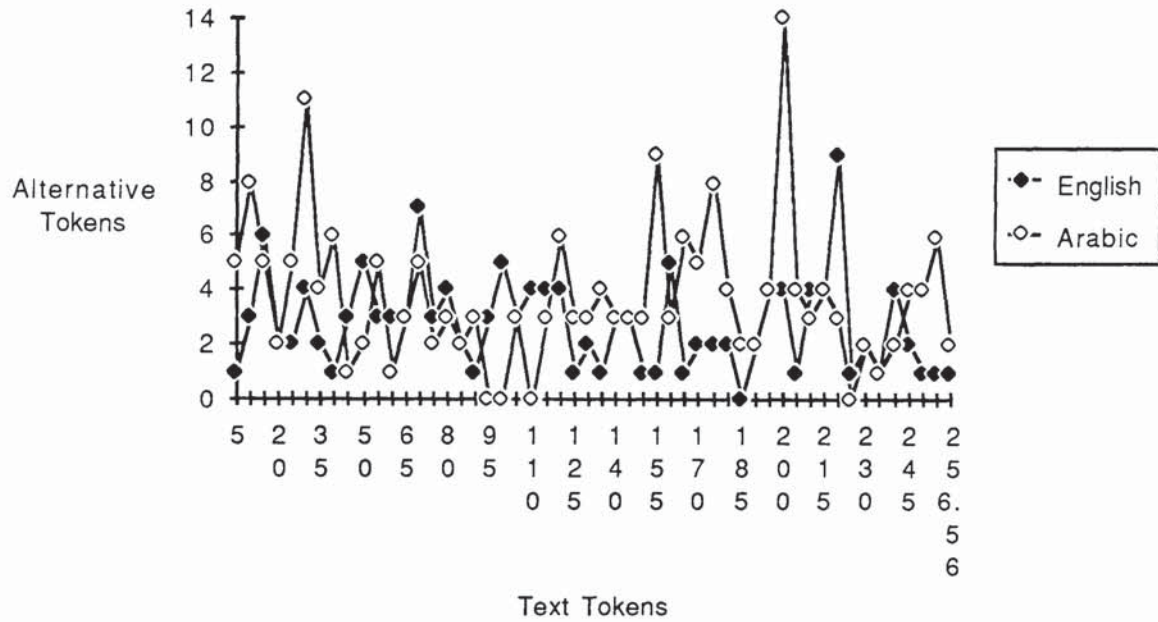
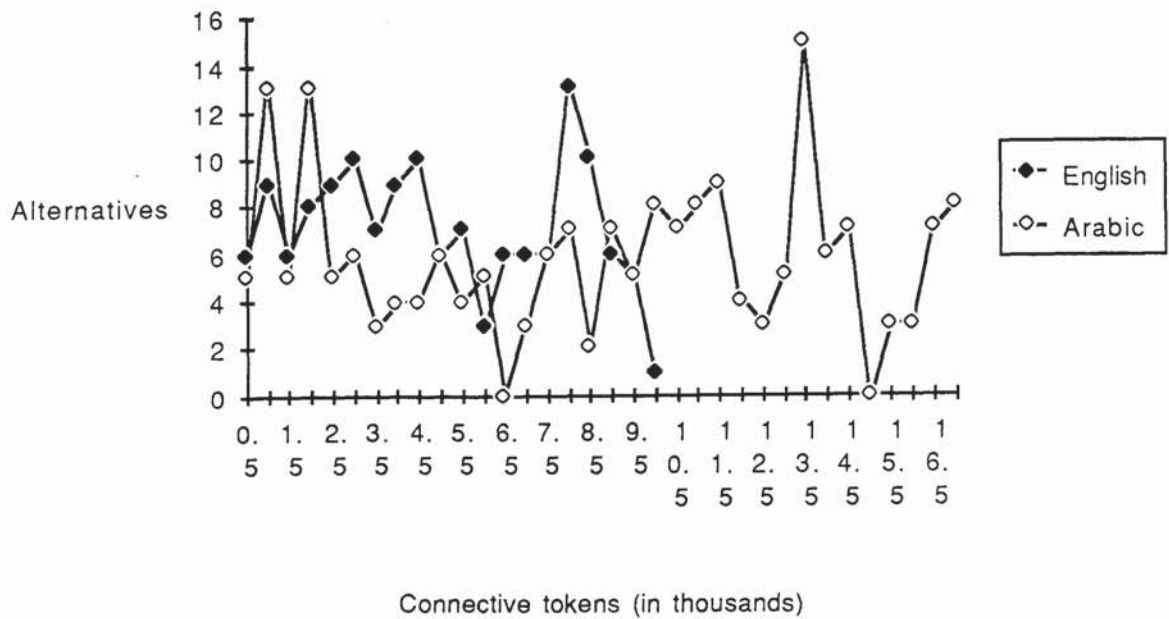


Fig. 9.27 Comparison of the Local Growth of Alternative Connectives



connective tokens is more extensive in English than Arabic. This is partly due to the fewer intervals in the English corpus (19 plus a shorter interval) compared to Arabic (34), which affects, to some extent, results of extrapolation. Accordingly an English corpus containing 100,000 connective tokens is expected to include 1,687 while an Arabic counterpart is estimated to have 983.

9.2.4.2 Textual Variations of Alternative Connectives

Observation of connectives of alternative relations in both corpora reveals a number of variations in textual behaviour. We shall discuss three: one is connective specific, while the other two concern the nature of the alternative relations, and are therefore related to the environment of the connectives.

1. Variation in Functional Coverage

There are two connectives in Arabic that collectively act as a counterpart to "or". Those are "'aw" and "'am". The first is the general connective of alternative relations in Arabic. The second, as a quick inspection of the relevant concordance will immediately point out, has a specific role: it combines two (yes/no) questions of the type in the following excerpt:

- [9.7] wa yuṣbiḥu al-su'ālu maṣrū'an. maḍā tastaṭī'u
'amrika 'an taf'ala? 'am hiya <ājizātun <an al-
ta'tīri?

[The question is legitimate. What can the American Administration do [in these circumstances]? Does it approve what is happening? Or [alternatively] is it incapable of exercising any influence?]

(Nb, 5/4/83, X4, 563-5)

The alternative relation is of the "exclusive/necessary" type (cf. 6.6 above) where only one option is true in a possible textual world. However, the way the questions are posed indicates that either a) the text producer is unaware which option is true and therefore leaves the choice open, or b) the question is rhetorical: the text producer is probably conscious of the answer but would like to influence the text receiver into formulating the same choice. In either way, the connective "'am" assists in establishing the options.

In English, however, this particular function does not need to be differentiated from the general patterning of the alternative function signalled by "or". Hence it does not require a separate connective, as this excerpt illustrates.

[9.8] Can one tackle Kant again? Or should it be heroic
assault on the mist-covered battlements of Structuralism?

(Ti, 28/5/83, X208, 33028-30)

Accordingly, "or" has a wider functional coverage than "'aw".

2. Functional Gradation

One feature that is more peculiar to Arabic than English is the extent of fuzziness in the semantic scope of the alternative relation, particularly the "accidental" type (cf. 6.6 above). According to the conceptual nature of the two (or more) propositions that connectives combine, the alternative relation in both corpora reflects a semantic continuum (Figure 9.28). In both languages, one end of the continuum represents configurations that are conceptually independent from one another though compatible in the text world.

"restatement"; that is the subsequent proposition reformulates in an approximate, and most probably unintentional, way, the content of the antecedent; and b) the alternatives reflect semantic as well as structural parallelism, thus approximating a relation of "coupling". The following two excerpts exemplify respectively these two types of blend:

- [9.9] 'ānaḡāka tuṣbiḡu mu<ālaḡatu al-<amali al-'ibdā<iyyi min jānibayni, 'aw nanḡuru 'ilay-hi min wijhatayni...
 [Then the artistic work can be considered from two point, or considered from two perspectives...]
 (J, 22/2/83, X86, 16810-12)

- [9.10] mu'tamarātun ... tan<aḡidu wa tantahī fī hudū'in wa bi-dūni 'an tuṡīra ḡajḡatan 'aw yakūna li-ba<ḡi-hā 'ayyatu 'aṡḡā'a ...
 [Conferences ... are held and concluded in quiet, without creating any clamour or having any echoes...]
 (Hr, 23/5/83, X57, 10436-41)

9.2.5 Reformulatory Connectives

9.2.5.1 Quantitative Variations of Reformulatory Connectives

1. Frequency Distribution

Differences in the distribution of reformulatory connectives across both corpora are displayed in Figures (9.29-30). The first figure shows their distribution in relation to text tokens, i.e. the entire corpus, while the second monitors their distribution within connective tokens.

a. Arabic reformulatory connectives constitute a larger category both in terms of tokens and types than English. The number of Arabic tokens is 239, more than twice as many as its English counterpart (which is 115) and the number of types (27) is 1.5

Fig. 9.29 Comparison of the Distribution of Reformulatory Connectives

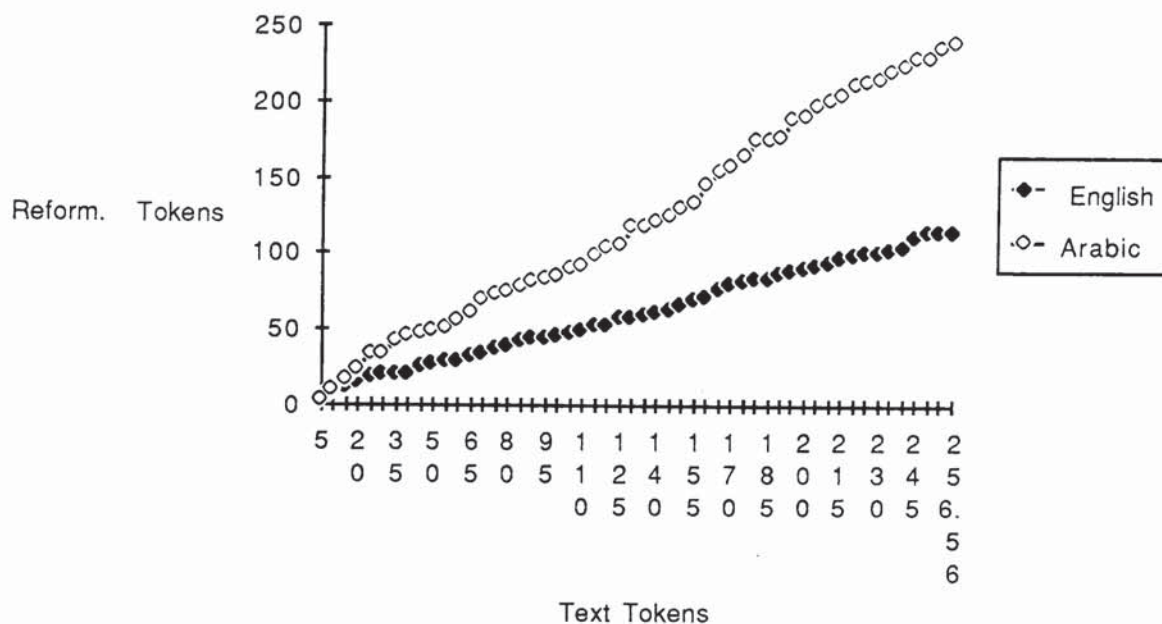
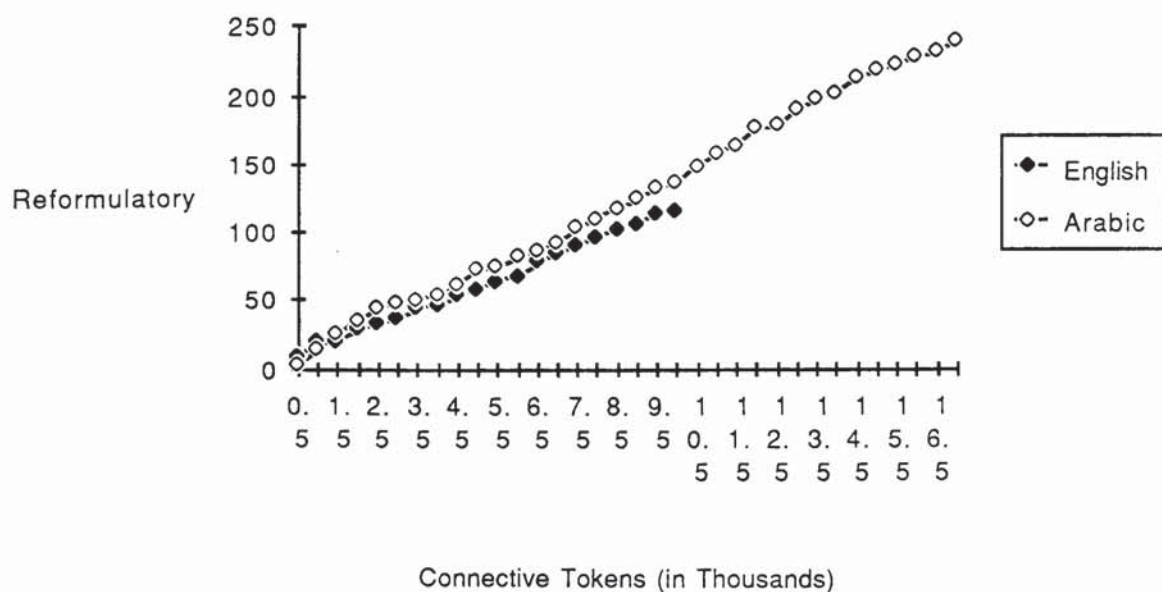


Fig. 9.30 Comparison of the Distribution of Reformulatory Connectives within Connective Tokens



larger. However, in terms of relative frequency, the category represents 1.4% of token mass in Arabic and 1.2% in English. This is due to the fact that connectives in Arabic comprise a larger set than in English.

b. In both corpora, the most frequent connective dominates the token shares: 33.5% in Arabic and 31% in English. In addition, the combined frequencies of the top 5 connectives take nearly similar share of token mass: 75% in Arabic and 72% in English. This means that the rest of the types in Arabic have smaller shares than in English.

c. Since there are 12 hapaxes in Arabic and 6 in English, the distribution points to a higher index of variegation in Arabic (44 compared to 33), higher density (8.293 compared to 7.223), rhythmicality (14.13 compared to 8.08) and stereotypicality (15.13 compared to 9.08). Concentration index is lower in Arabic: 0.104 compared to 0.062. Indices of gravity and exclusivity of this category manifest some similarity in both corpora. They are respectively 0.050 in Arabic vs. 0.052 in English, and 20 vs. 19 in English.

2. Repetitiveness

Due to the relatively bigger size of this category in Arabic, connectives of reformulation manifest more repetitiveness within the Arabic than English corpus. This finding is characteristic of Arabic and is indicative of certain strategies in textual organisation.

Comparison of indices of repetitiveness give the following results:

a. Connective occurrence rate is higher in Arabic than in English (83 compared to 71).

b. The repeat rates are also higher in Arabic than in English. The general repeat rate is (20×10^{-5}) in Arabic and (14×10^{-5}) in English. The reformulatory system repeat rate is (33×10^{-6}) in Arabic and (22×10^{-6}) in English.

c. The gap distribution displays relatively shorter distances separating two occurrences of reformulatory tokens in Arabic than in English. The average distance is 1,067 words in Arabic and 2,209 in English.

d. Type occurrence rate is the only index of repetitiveness that is higher in English in comparison to Arabic (6 compared to 9). This is due to the fact that the proportion of Arabic to English tokens is higher than the proportion of types.

3. Growth

Comparisons of the global and local growth are displayed in Figures (9.31-32) respectively. A summary is offered below:

a. Global Growth

i. Inspection of growth of reformulatory tokens in both corpora shows faster growth of tokens during the early intervals in English than in Arabic. At interval 10 (50,000 words) English reformulatory tokens achieve 23% of reformulatory token coverage while in Arabic

Fig. 9.31 Comparison of "Global" Growth of Reformulatory Connectives

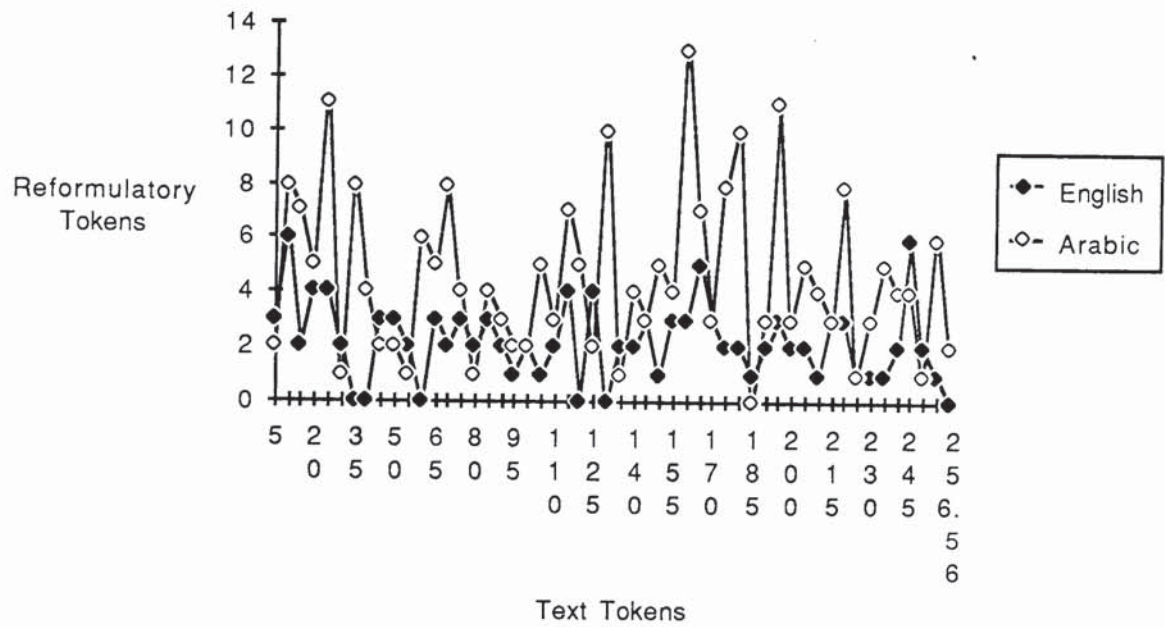
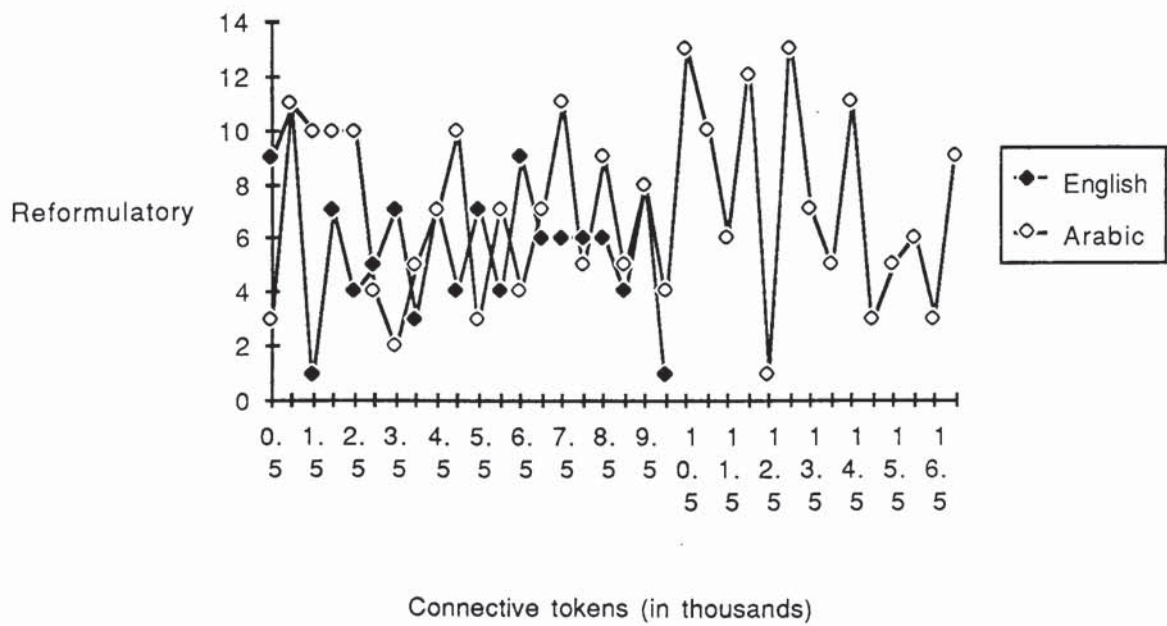


Fig. 9.32 Comparison of Local Growth of Reformulatory Connectives



it achieves 20%. The next 10 intervals achieve 17% of token mass in English and 16% in Arabic. However, growth rate after that decreases in the English corpus and gradually increases in the Arabic corpus. For instance, the third 10 intervals achieve 19% of tokens compared to 17% in English. The fourth 10 intervals show a growth of 25% of tokens in Arabic and 21% in English.

This behaviour in growth influences results of extrapolation. For instance, a 500,000 word Arabic corpus is expected to contain 478 tokens in Arabic; an English corpus of the same size is estimated to contain 196. A million word Arabic corpus is expected to contain 963 tokens while an English corpus of a similar size is expected to contain 358.

ii. Reformulatory types show relatively faster growth in the English corpus until the corpus mid-point, where 78% of types will have appeared. At the same point in the Arabic corpus, 67% of types will have been achieved. After this point growth decreases to a lower rate than Arabic. This is an indication of a recycling of types within English, which is commensurate with the general profile of growth of English connectives.

Results of extrapolation show that an Arabic corpus of a million word length is expected to contain 53 types while an English corpus of a similar size is expected to contain 31.

b. Local Growth

The relatively bigger size of this category in Arabic is reflected in the calculation of Tuldava's richness index, which in

this case computes the size of reformulatory connectives within successive text intervals each containing 500 connective tokens. The index is noticeably higher in Arabic: 5.0599 compared to 2.0347 in English. In contrast, due to the larger size of connective mass within the Arabic corpus, which means that the corpus is divided into more intervals than English, the growth rate index within the intervals is lower in Arabic: -4.0977 compared to -3.03188 in English. These indices influence the results of extrapolation. A corpus that contains 100,000 connectives in Arabic is expected to comprise 1,614 reformulatory tokens. An English corpus that contains a similar number of connectives is expected to include 824.

4. Distribution of Reformulatory Categories

Comparison of the distribution of reformulatory categories in English and Arabic is displayed in Figures (9.33-34). The following statements give a brief summary.

a. Restatement

This category is larger, both in terms of tokens and types, in the Arabic corpus than English. The absolute frequency of the category is 136 tokens and 15 types in Arabic compared to 50 tokens and 12 types in English. In terms of relative frequency the category represents 57% of reformulatory connectives in Arabic and 43% in English.

In Arabic, the category is dominated by two connectives, the combined frequency of which represents 73% of restatement tokens. These connectives are "ay" [that is] and compound connectives where

Fig. 9.33 Comparison of the Distribution of the categories of Reformulatory connectives in the two Corpora

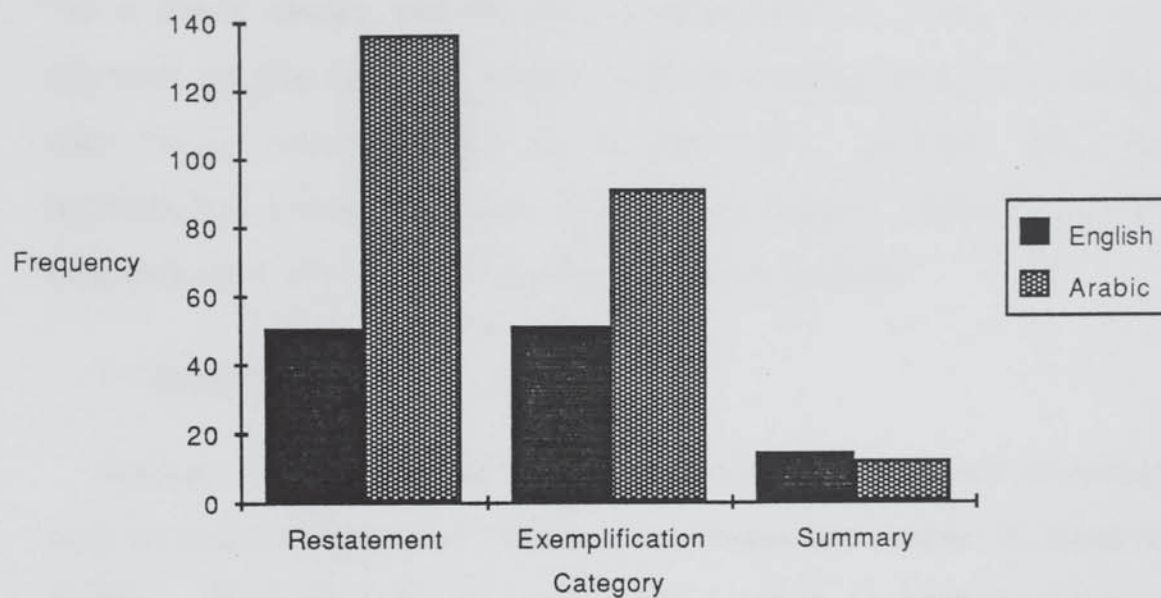
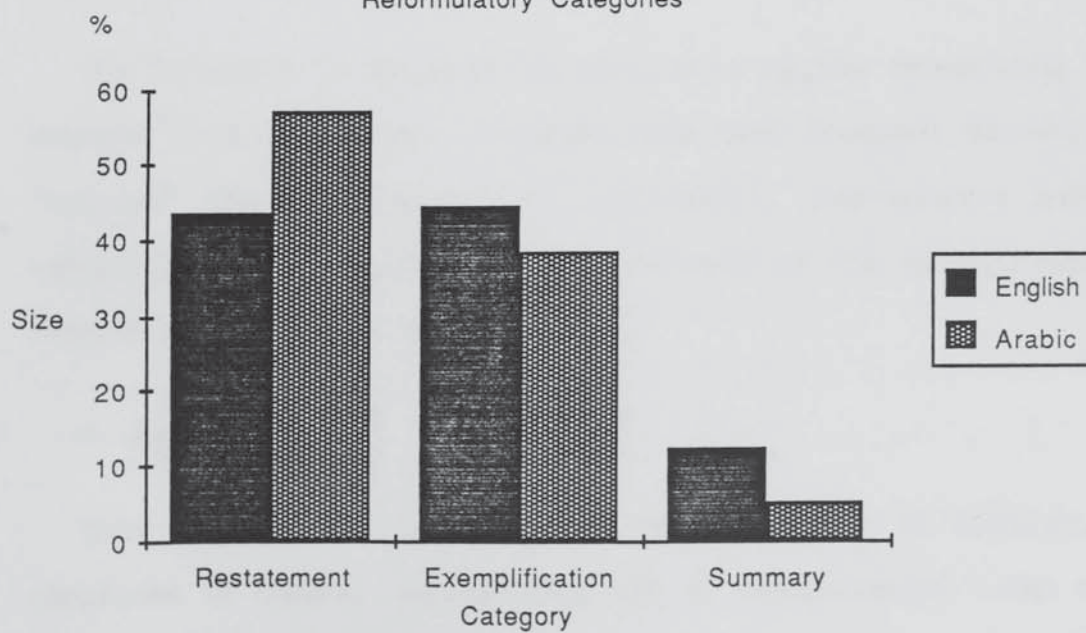


Fig. 9.34 Comparison of the Distribution of Shares of Reformulatory Categories



"ma<nā" is a core: "ma<nā dālika 'anna" and "bi-ma<nā 'anna" [meaning, this means]. In contrast, a similar share in tokens in English is achieved by 4 connectives: "in other words", "rather", "in a (real) sense", and "to put (it differently)". This means that the rest of the types in Arabic receive smaller shares in tokens than their counterparts do in English. Indeed, the rank distribution shows 7 hapaxes (47% of restatement types) in Arabic compared to 4 (33% of restatement types) in English.

b. Exemplification

Similar to the previous category, this one is larger in Arabic both in terms of absolute frequency of tokens and number of types in Arabic. It comprises 91 tokens and 6 types in Arabic while it contains 51 tokens and 3 types in English. However, their shares in reformulatory token mass is different. In Arabic the category represents 38% of tokens while in English it stands for 44%.

The category in English is dominated by the connective "for example" (71% of tokens). In Arabic the most frequent connective, "maṭalan" [for example, for instance], represents 44% of exemplification tokens, allowing the rest of the connectives to receive larger shares in tokens.

b. Summary

This is a small category in both corpora. In English, it comprises 14 tokens, representing 12% of reformulatory token mass, while in Arabic it contains 12 tokens, equal to 5%. In terms of types, each corpus contains 4. In English, the category is

dominated by "in short" (79% of summary tokens) while in Arabic the dominant connective is "bi-'ixtiṣārīn" [in short, in brief] (75% of summary tokens).

9.2.5.2 Textual Variations of Reformulatory Connectives

Due to the small size of the subcategory of summary in both corpora, textual variations will only be discussed in the textual functioning and patterning of "restatement" and "exemplification".

9.2.5.2.1 Restatement

The quantitative account in the previous section points to a tendency, stronger in Arabic than in English, to uphold text development and make a regression to redefine, clarify or rephrase the content of a previous proposition. This is evident both in the frequency of the operation and the number of connectives that signal it.

To explain this further, we must note that in the process of text actualisation the text producer arranges the content of a set of propositions in a network of concepts and relations, and finds a surface expression that captures its conceptual organisation in a text world. In performing this task the text producer may have a good reason to be dissatisfied with his network and particularly the way the concepts are mapped. To rectify any source of deficiency, he halts the process by generating an alternative version of the network. This new version will adopt a more flexible standpoint regarding concept boundaries, interestingness, and/or clarity. This

tendency is more powerful in Arabic rhetorical organisation than in English.

9.2.5.2.2 Exemplification

The quantitative contrastive profile made above underlies another interlingual variation in the way connectivity assists textual organisation. Rhetorically, the use of exemplification responds to a pressure within text for illustrating a detail, substantiating a view or an argument and providing a persuasive proof. This response is more readily achieved by connectives in Arabic than in English.

A further variation involves the organisational nature of the pattern of exemplification that the connectives are associated with. Two variants of this pattern have been observed in the two corpora, which we shall label Pattern A and Pattern B. Again, we have to express this variation, though qualitative in essence, in quantitative terms. That is, there is a stronger tendency - in other words, higher probability - for one pattern to appear in one language than in the other.

Pattern A is the basic textual pattern of exemplification. A statement is made that expresses a generalisation in the textual world and may take the form of a fact, viewpoint, an observation, etc. The statement is expanded by exemplification, that is by reformatting part of its conceptual content. This is assisted by the appropriate connective which explicitly signals the relation (as well as the underlying conceptual link). The relation falls into one of the two functional patterns discussed in 6.7.3. The subsequent, i.e. the exemplificatory statement, represents a self-

contained block of knowledge the role of which is exhausted once the relation is established. This allows the text producer to resume developing his text either by making a statement related to the generalisation or initiating a new text sequence. This pattern of exemplification is the norm in the English corpus. For example in this excerpt

[9.11] ...though there has never been a challenge to the Irish law, the supreme Court [in Ireland] has not been afraid to tinker with other statutes. For example, contraception is, by and large, now legal in the Republic.

It is a rare politician in Ireland who dares to take issue with Roman Catholic dogma, ...

(O, 23/1/83, X49, 7727-34)

the exemplificatory statement, introduced by "for example", is self-contained in the sense that it does not require further expansion.

In Pattern B, however, the exemplificatory statement is not self-contained. It exercises a textual influence on later sequencing and therefore plays an active role in directing the development of text. This is usually achieved when the exemplificatory statement creates sufficient salience that a shift in focus takes place whereby the content of the antecedent (the statement of generalisation) is informationally/conceptually subdued. This pattern is more dominant in the Arabic than English corpus. In the following excerpt the writer makes a generalisation stating that there are countries in Asia with similar economical problems to those of Egypt and yet managed to overcome most of them by creativity and hard work. Exemplification is offered in the ways that Thailand has faced its problems and the next few paragraphs in the text expands this statement.

[9.12] ...lā yastaṭī'ū al-'insānu ... 'an yansā laḥẓatan waḥīdatan bilāda-hu ... fī daw'i mā yarā-hu min 'i<jāzin baṣariyyin li-ṣu<ūbin wājahat zurūfan muṣābihatan fī ba<ḍi jawānibi-hā li-zurūfi bilādi-nā, lākinna-hā tagallabat <alay-hā bi-al-'irādati al-ḥadīdiyyati ...

Fa al-taqaddumu wa al-numuwwu wāḍihāni maṭalan fi Tayland bi-ṣūratin tad<ū li-al-dahṣati faḍlan <an al-'i<jābi. fa fī kulli buq<atin min-hā binā'un wa taṣyīdun, wa siyāḥatun wa <amalun wa jiddiyyatun min al-ṣigāri wa al-kibāri <alā al-sawā'i.

[One cannot forget, even for one moment, one's own country as one encounters human miracles in countries that have faced problems similar in certain respects to ours, yet have overcome them with iron will ...

For instance, the progress and development in Thailand are sources of wonder and admiration. For in every spot in the country there is planning and construction, tourism and industry, agriculture and export, serious work and perseverance both among the younger and older generations.]

(Hr, 1/4/83, X45, 8186-8201)

In other instances in the corpus this pattern may undergo some textual variation, but it basically remains the same overall pattern. The variations only affects the size of the influence it exerts on creating and organising later sequences of text.

9.2.6 Orientative Connectives

9.2.6.1 Quantitative Variations of Orientative Connectives

1. Frequency Distribution

a. English orientative connectives constitute a larger category, both in terms of size of token mass and type repertoire, than that of Arabic orientatives (see Figures 9.35-36). It comprises 1,011 tokens and 127 types while its counterpart in Arabic contains 738 tokens and 92 types. The difference is intensified even further

Fig. 9.35 Comparison of the Distribution of Orientative Connectives

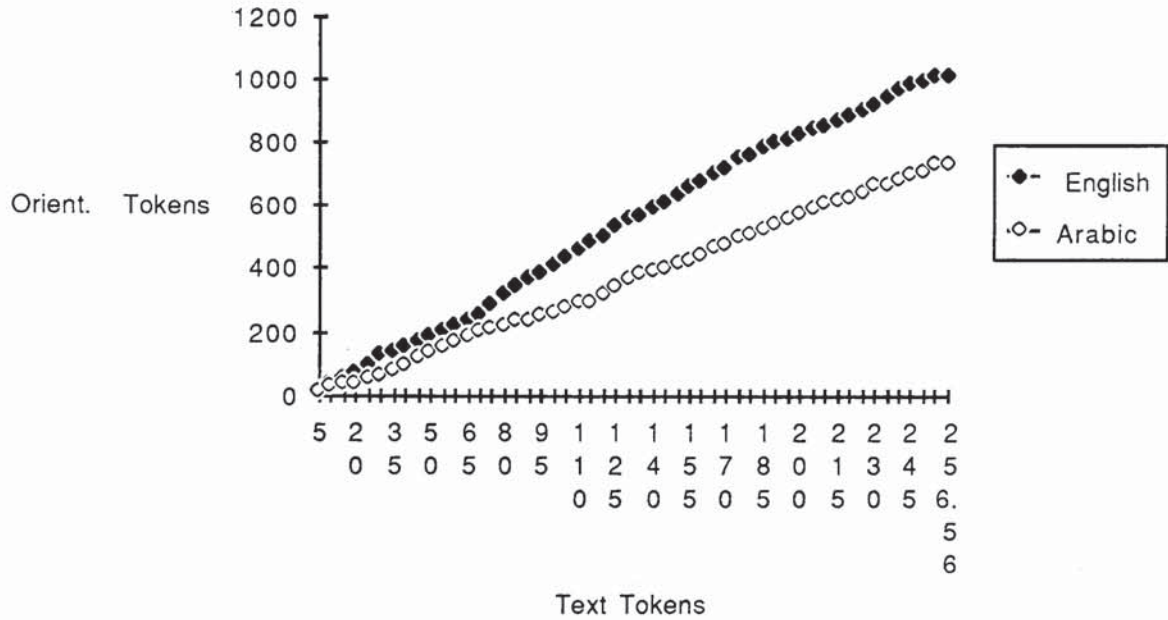
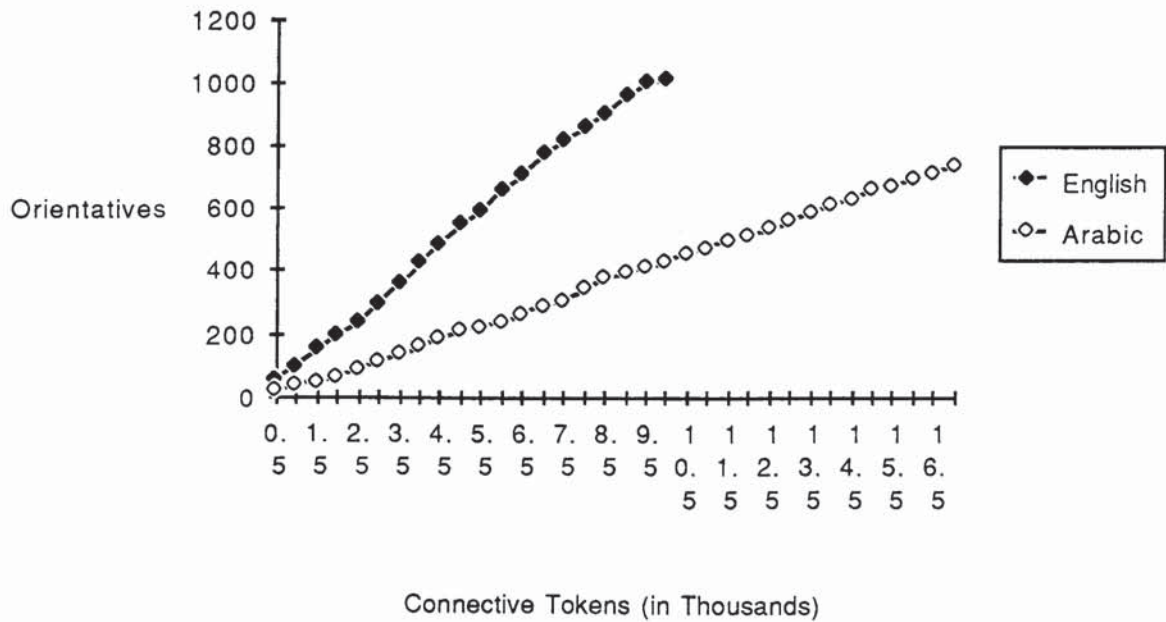


Fig. 9.36 Comparison of the Distribution of Orientative Connectives within Connective Tokens



when we consider the shares that each category has within connective token mass: English orientatives represent 10.5% while Arabic orientatives stand for 4%. This finding is indicative of the importance of this category in organising text components and signalling relations in English.

b. Inspection of the frequency distribution reveals that fewer types achieve 75% of orientative token coverage in Arabic than in English: 19 compared to 27. This indicates fewer connectives dominating the orientative token mass in Arabic.

c. Despite the difference in the size of the category across the both corpora, their TTR is nearly similar: 0.126 in English and 0.125 in Arabic. This is indicative of similar richness and extensity of use of types.

d. English orientatives comprise 47 hapaxes whereas Arabic contains 38. But because of the bigger number of types in English, the variegation index is larger in Arabic than English (41 compared to 37) and the concentration is lower (0.073 vs. 0.079). Gravity of hapaxes is lower in English, due to the larger number of orientative tokens, the indices being 21 compared to 19 (i.e. more hapaxes tend to appear in shorter distances in Arabic). Indices of rhythmicality, stereotypicality and density are close in both corpora: respectively 11 vs. 12 in Arabic, 12 vs. 13, and 10.47 vs. 10.04.

2. Repetitiveness

In general, English orientatives manifest higher rates of

repetitiveness within the corpus than Arabic, which reinforces the finding made in (1) above on the significant textual role that this category has connectivity in English. The indices that have been compared are as follows:

a. Orientative occurrence repeat rate is higher in English (17 compared to 23). Type occurrence rate is similar in both corpora, calculated at 8.

b. The general repeat rate and orientative system repeat rate are both higher in English (respectively 0.011 compared to 0.002 and 67×10^{-5} compared to 12×10^{-5}).

c. The gap distribution of orientative reveal shorter distances separating occurrences of connectives in English. The average distance in English is 252 words and in Arabic 346.

3. Growth

Comparisons of global and local growth are displayed in Figures (9.37-38) respectively. The following statements give a brief account.

a. Global Growth

i. Calculation of growth within each corpus manifests similarity in growth rate of tokens within the various intervals. The index is calculated, using Tuldava's formula, as -4.015065 in English and -4.028952 in Arabic. Richness index, indicating here the size of the growth within each interval in relation to the size of the interval, is also similar in both corpora: 1.048117 in English and 1.073594 in

Fig. 9.37 Comparison of "Global" Growth of Orientative Connectives

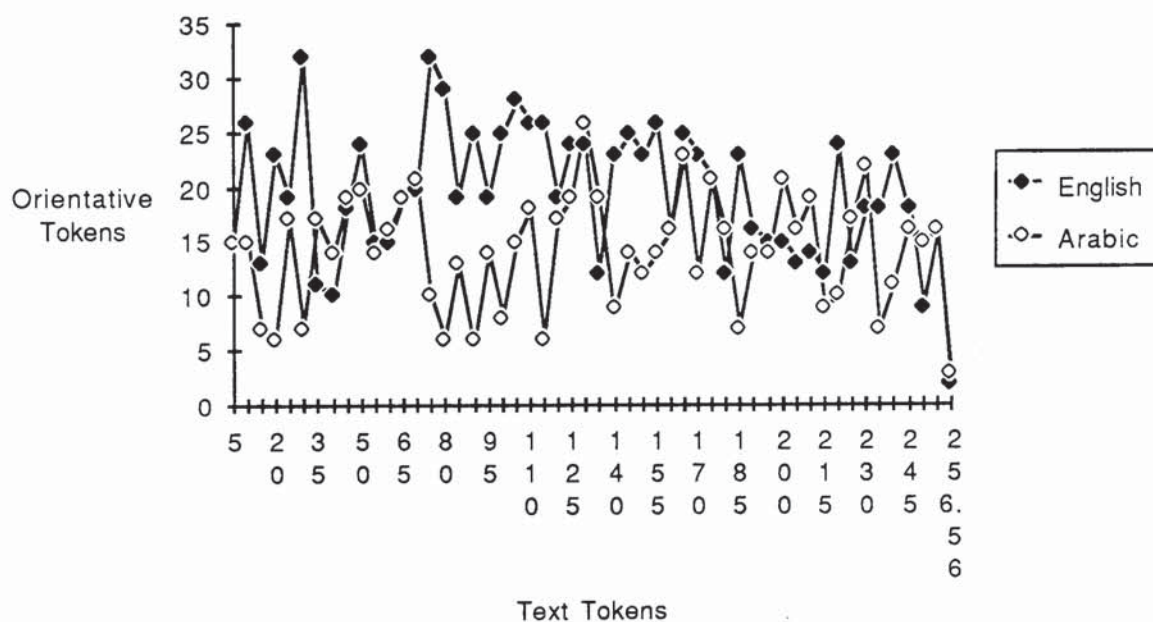
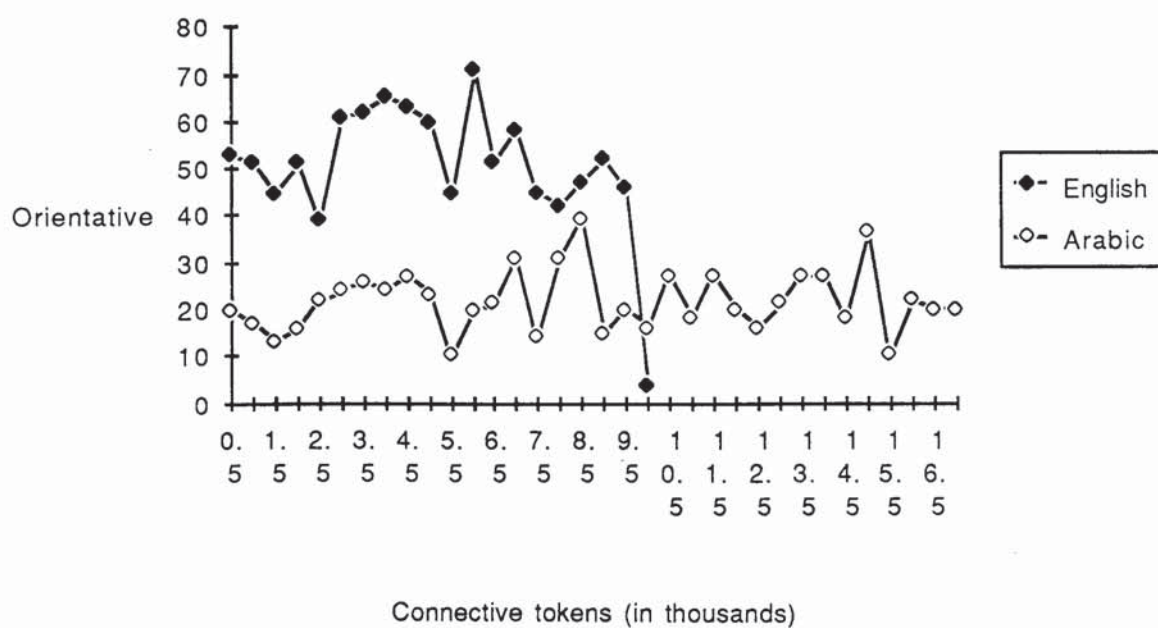


Fig. 9.38 Comparison of the Local Growth of Orientative Connectives



Arabic. A million word corpus is expected to contain 4,417 orientative tokens in English and 2,959 in Arabic.

ii. Growth of types shows some variation. Although the 50th percentile type occurs within interval 12 (60,000 words) in both corpora, the 90th percentile type occurs within interval 38 (190,000 words) in English and 34 (170,000 words) in Arabic. This shows earlier saturation of orientative types in Arabic than English. This finding confirms that orientatives constitute a more open class in English than Arabic.

b. Local Growth

Due to the smaller total number of connective tokens in English compared to Arabic, growth rate of orientatives within text intervals each containing 500 connective tokens is greater. It is computed, using Tuldava's formula, as -2.047613 , whereas Arabic growth rate is calculated as -3.096516 . Results of extrapolation show that an English corpus containing 100,000 connective tokens is expected to include 11,246 orientative tokens, whereas an Arabic corpus having the same number of connective tokens is expected to contain 4,798.

4. Distribution of Orientative Categories

Variations in the distribution of orientative categories across the two corpora are displayed in Figures (9.39-40). Below is a brief account for each category.

Fig. 9.39 Comparison of the Distribution of the Categories of Orientative Connectives in the two Corpora

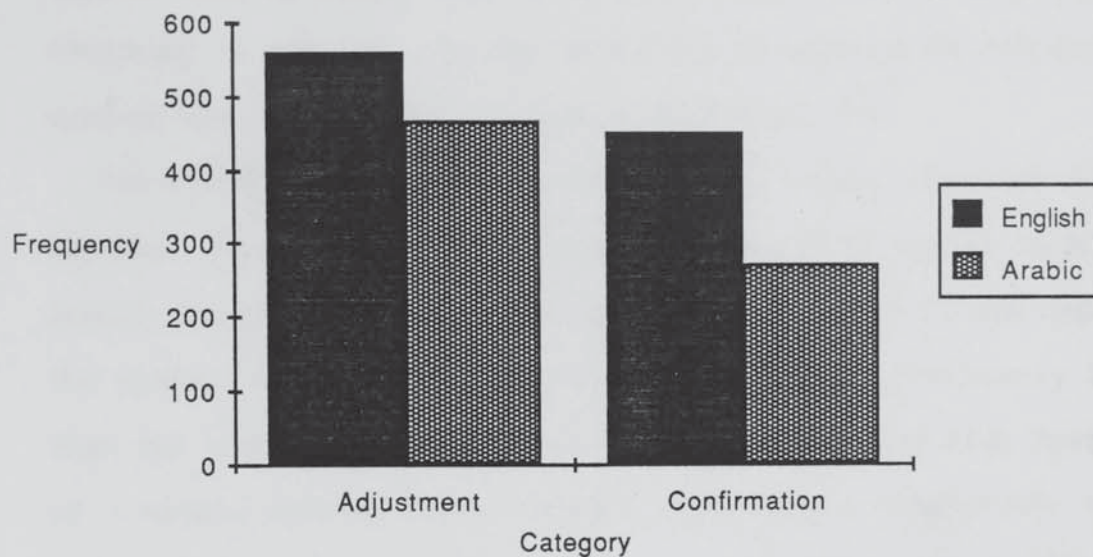
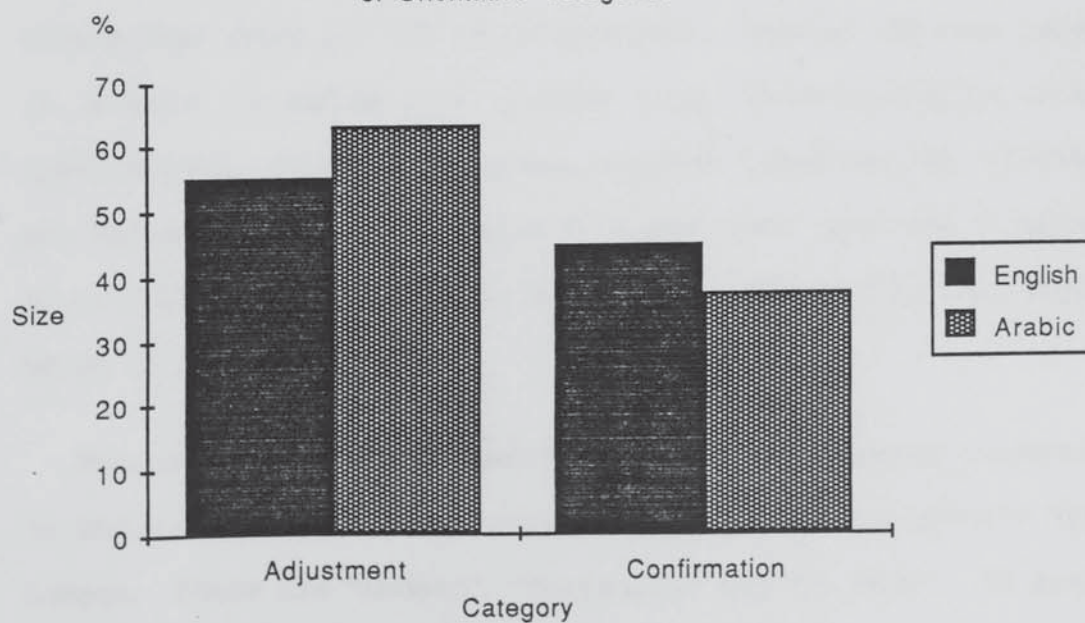


Fig. 9.40 Comparison of the Distribution of Shares of Orientative Categories



a. Adjustment

Although the absolute frequency of this category is higher in English than in Arabic (559 connective tokens vs. 465), its relative frequency is smaller. It represents 55% of orientative connectives, whereas the same category in Arabic stands for 63%.

The category in English comprises 103 types, of which 42 are hapaxes. The number of types that achieve 75% of tokens is 28. In Arabic the category comprises 66 types, of which 31 are hapaxes. The number of types achieving 75% tokens is 12, distinctly lower than its counterpart in English. This is due to the high frequency of 3 Arabic connectives: "'ammā", "kamā", and "xāṣṣatan", whose combined frequency represents 54% of connectives of adjustment.

b. Confirmation

The size of this category is larger in English both in terms of absolute and relative frequencies than in Arabic. It comprises 452 tokens that stand for 45% of orientatives, whereas the same category in Arabic contains 273 tokens that correspond to 37% of orientatives. In terms of types, English comprises 26, of which 6 are hapaxes. Arabic includes 27 types that comprise 7 hapaxes. Accordingly, TTR is higher in Arabic, computed as 0.1, than English, where it is 0.06.

Both categories are influenced by the high frequency connectives. In English the top 3 connectives represent approximately 50% of tokens. These are "indeed", "certainly" and "in fact". In Arabic, the same size of token coverage is achieved by 4 connectives: "lā šakka", "fi<lan", "'innamā" and "bi-al-ṭab<i".

9.2.6.2 Textual Variations of Orientative Connectives

9.2.6.2.1 Adjustment

In order to understand the interlingual variation in the textual utilisation of this category of orientative connectives, we have to view it from three related perspectives: diversity of signalling devices, mobility of connectives and the phenomenon of textual detachment in Arabic.

1. Diversity of signalling devices

The quantitative variation described above manifests a larger and more varied number of devices for signalling "adjustment" in English than in Arabic. This diversity of devices in English is indicative of a pressure to clarify textual organisation by signposting various transitions within a text. These transitions are usually of two types:

a. Major transition: This designates a transition between two large text "chunks" each comprising at least a paragraph and may in some cases run into several paragraphs.

b. Minor transition: This is a transition that occurs between two small text sequences within the paragraph boundary.

The transition usually implies a kind of topical shift, either slight or sharp, in the conceptual content of the next text sequence from the current one. This involves operations of regression and progression whereby activities of processing are directed towards

some specified goal nodes. The connectives act as a switch whereby the processor is helped to make the transfer from one track to another smooth, even and above all, accurate.

Rhetorically, the indication of topical shift enables the text writer to step directly into developing the transition. In semantic/rhetorical terms, transitions constitute diverse varieties, depending on how the intended conceptual shift adjust the rhetorical development of the text or text sequence. Although we have grouped them into 6 functional categories (see 6.8 above), each comprises a detailed spectrum of semantic variations. Since there is a strong pressure in English to signal them, and since sharpness and accuracy form a necessary requirement for signalling transitions, it follows that a great diversity of signalling devices (i.e. connectives) are available to be utilised in text development.

In Arabic, as inspection of the corpus confirms, pressure for signalling devices exists, but it is less powerful and less urgent than in English. Unless the transition is sufficiently acute to justify overt signalling, it is either left unmarked and has to be inferred through the mere juxtaposition of the two text sequences, or diminished through additivity, i.e. use of "wa" to combine the sequences. In some cases, although the transition is made explicit by a connective that adjusts the perspective of the subsequent, it is slackened via additivity. That is, an additive connective, usually "wa", precedes the connective of adjustment and reduces the impact of the semantic shift that the second sequence carries.

2. Mobility of Connectives

Structurally, English connectives of this category are adverbials, mostly disjuncts (for a detailed treatment of adverbials in English, see Greenbaum 1969, Quirk et al. 1972, 1985, Quirk and Greenbaum 1973). Since adverbials are mobile, that is they permit variability of positioning, they give the text producer the choice to signal any transition directly by having the connective at the beginning of the sequence (which, incidentally, is the most frequent position in the corpus). Alternatively, the signalling can be delayed and the connective is inserted within the textual sequence. In either case the decision is a rhetorical one and the text producer chooses the positions where the connective can best maintain the cohesiveness of the sequences. For instance, compare the position of "in particular" in these two excerpts:

- [9.13] ...the higher interest rates which the US Treasury has to pay to finance its big budget deficits (estimated at \$200 billion for 1984) pull funds into the dollar. In particular, the dollar's latest surge forward has been based on the development of record high real interest rates.

(G, 19/11/82, X3, 419-25)

- [9.14] There were already two elder brothers, Wilfred and Christopher, and it was the eldest in particular who enthusiastically guided the first steps of Anthony Blunt's progress through the galleries of that extraordinary Parisian scene.

(G, 28/3/83, X18, 2753-7)

More examples of the variable positioning of other connectives can be inspected in the relevant concordance.

In contrast, Arabic connectives of adjustment show, generally

speaking, little variability of position. The normal position is the head of the subsequent sequence, i.e. the one that carries the topical shift, whether it is a clause, sentence or a paragraph. The few connectives that manifest more mobility than others are those that signal specification of perspective, particularly "fī hādhi al-ḥālātī", "fī hādā al-sadādī" and "fī hādā al-majāli" [in this case, respect].

3. Problem of Detachment

The use of some Arabic connectives in initial position (even when preceded by "wa") in the subsequent sequence, and in particular the connective "'ammā", creates two contradictory roles. On the one hand they have the role of sustaining text cohesion by signalling text transition and thus marking adjustment of conceptual perspective. On the other, their use creates a gap that detaches it from the subsequent sequence, particularly when the structural pattern of the latter involves a nominal clause. In the case of "'ammā" a gap is present whatever structural pattern the subsequent clause may have. This discrepancy in textual role is removed by bridging the gap with "fa" (cf. 6.4.6 above). The following two excerpts exemplify this problem:

- [9.15] 'ammā taškīlu al-lajnatī fa lā qīmata la-hu.
[As for the formation of the committee, it has
no importance]
(Nb, 15/7/83, X22, 4787)

- [9.16] wa min hādā al-muntalaqi fa laysa tammata man
yataḥammalu al-mas'ūliyyata al-'ulā <an 'irtikābi
maḍbaḥātī muxayyamay ṣabrā wa šātīlā siwā al-
wilāyātu al-muttaḥidatu al-'amrikiyyatu.
[And within this perspective, the country that is
immediately responsible for the massacres of

In both excerpts, the connectives create a problem of detachment and sequencing is resumed via the use of "fa". This textual tendency does not associate the cohesive role of connectives of adjustment in English.

9.2.6.2.2 Confirmation

Confirmation represents one of the essential means of establishing authorial tone in text. It involves qualifying the text producer's stance or attitude towards the content of certain sequences in his text. This is accomplished when the text producer responds to his current statement and offers a positive qualification of the content. There are, of course, different structures that establish textual confirmation, one of them being the use of some adverbials in a cohesive role (cf. 6.8.3).

It is evident from the profile of quantitative interlingual variations within the two corpora that English resorts to this type of signalling far more frequently than Arabic. The connectives themselves set in motion a major mechanism for organising and developing text sequences whereby information, conceptual or perceptual, is not only ordered in a particular position in the text, but is maintained through approval and endorsement, and is therefore placed under focus. This organising mechanism is far more active in textual development in English than in Arabic.

9.2.7 Temporal Connectives

9.2.7.1 Quantitative Variations of Temporal Connectives

1. Frequency Distribution

a. In terms of absolute frequency, temporal connectives constitute a bigger category in Arabic than English. It comprises 1,494 tokens and 62 types, whereas the same category in English contains 1,284 tokens and 48 types. The TTR, indicative of richness and extensivity of use of types, is 0.041 in Arabic and 0.037 in English. In terms of relative frequency, however, this category represents 9% of total connectives in Arabic, whereas it constitutes 13% in English. Figures (9.41-42) display comparisons of the distribution of both categories within the corpora and in relation to connective tokens.

b. Inspection of the distribution shows similarity in the relative frequency of the top 5 connectives in both categories. In combination, they represent 61% of temporal token mass in each corpus.

c. Arabic temporal connectives comprise a bigger number of hapaxes, 19 compared to 9 in English. This affects a number of quantitative features in both categories. For while concentration indices are similar, both computed as 0.03, features such as variegation, exclusivity and gravity are different. Exclusivity is higher in Arabic, computed as 0.0127, whereas the index in English is 0.007. Variegation is also higher in Arabic: 30.6 compared to 10.7. Gravity is computed as 79 in Arabic and 143 in English.

Fig. 9.41 Comparison of the Distribution of Temporal Connectives

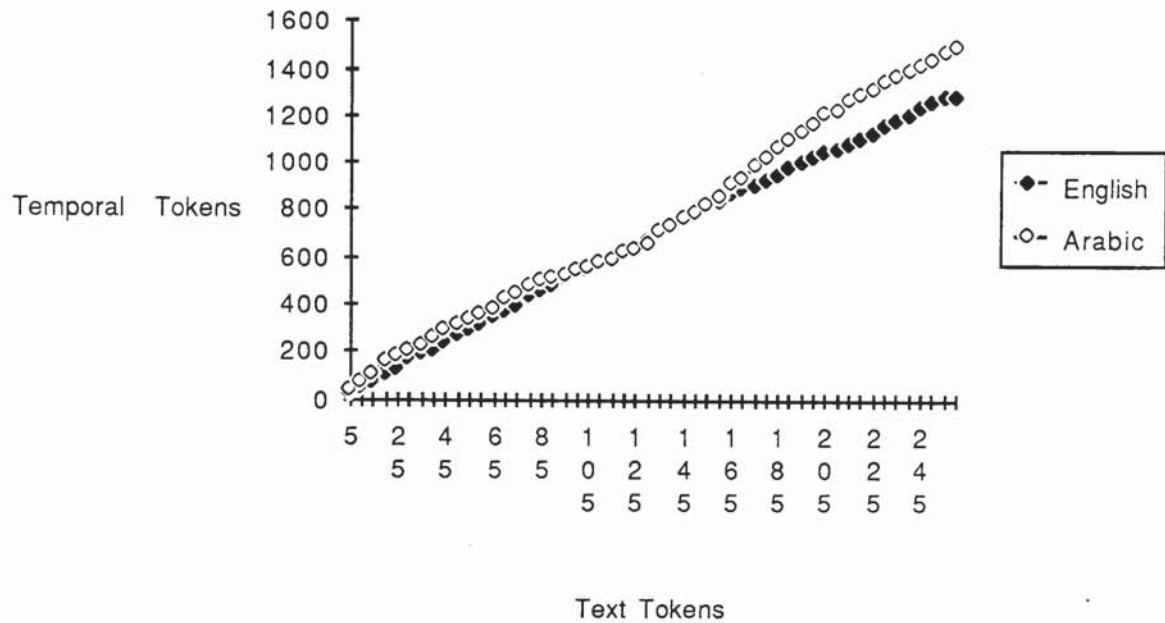
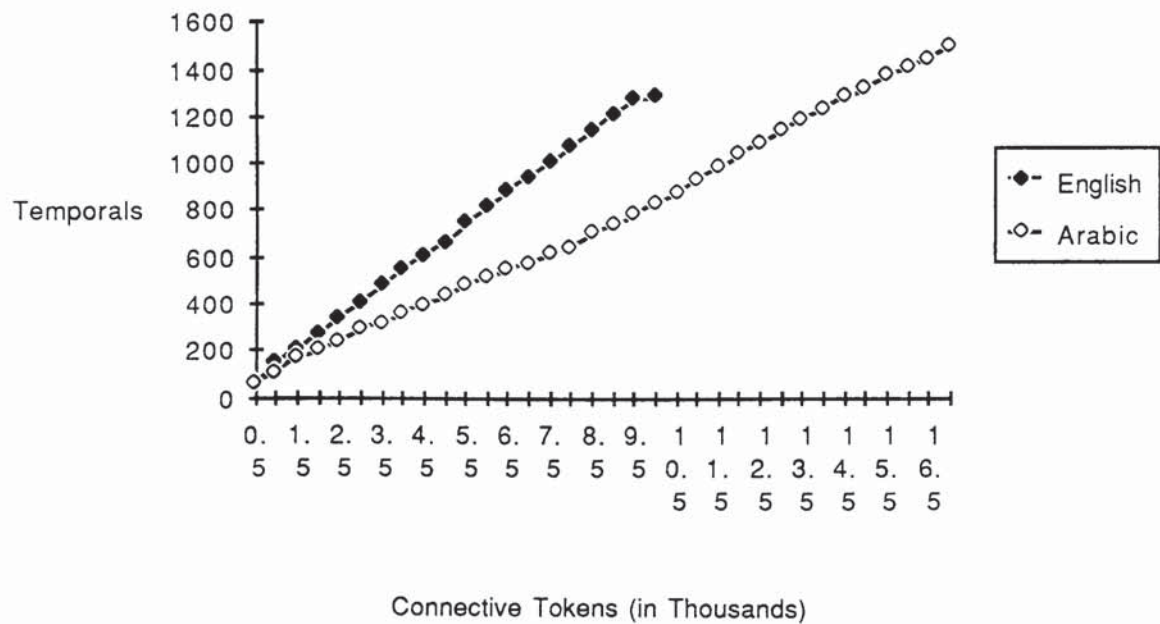


Fig. 9.42 Comparison of the Distribution of Temporal Connectives within Connective Tokens



Rhythmicality and stereotypicality are higher in Arabic: respectively 33.3 in Arabic vs. 31.69 in English and 34.3 vs. 32.69.

2. Repetitiveness

In general, temporal connectives in English manifest higher rates of repetitiveness than their counterparts in Arabic. This is evident in the higher temporal occurrence rate (computed as 7.5 compared to 11 in Arabic), type occurrence rate (27 vs. 28), general repeat rate (0.018 vs. 0.008) and system repeat rate (24×10^{-4} vs. 9.4×10^{-4}). However, due to lower numbers of tokens, average distance between two occurrences of temporal connectives is lower, computed as 199 in English and 171 in Arabic.

3. Growth

Comparisons of global and local growth of temporal connectives in both corpora are displayed in Figures (9.43-44) respectively. The following statements give a brief account.

a. Global Growth

i. Growth of tokens fluctuate within the various successive intervals in both corpora. But, generally, growth within the English corpus is more consistent. For instance, the first 10 intervals (representing approximately 0.2 of the corpus) contains 21% of temporal tokens. At mid-point growth achieves 50% of temporals. In the Arabic corpus, growth fluctuates more noticeably. At mid-point 46% of temporal tokens have been achieved. This fluctuation affects computing growth rate; the index is higher in Arabic than English (-3.0956 vs. -5.0121). It also affects results

Fig. 9.43 Comparison of "Global" Growth of Temporal Connectives

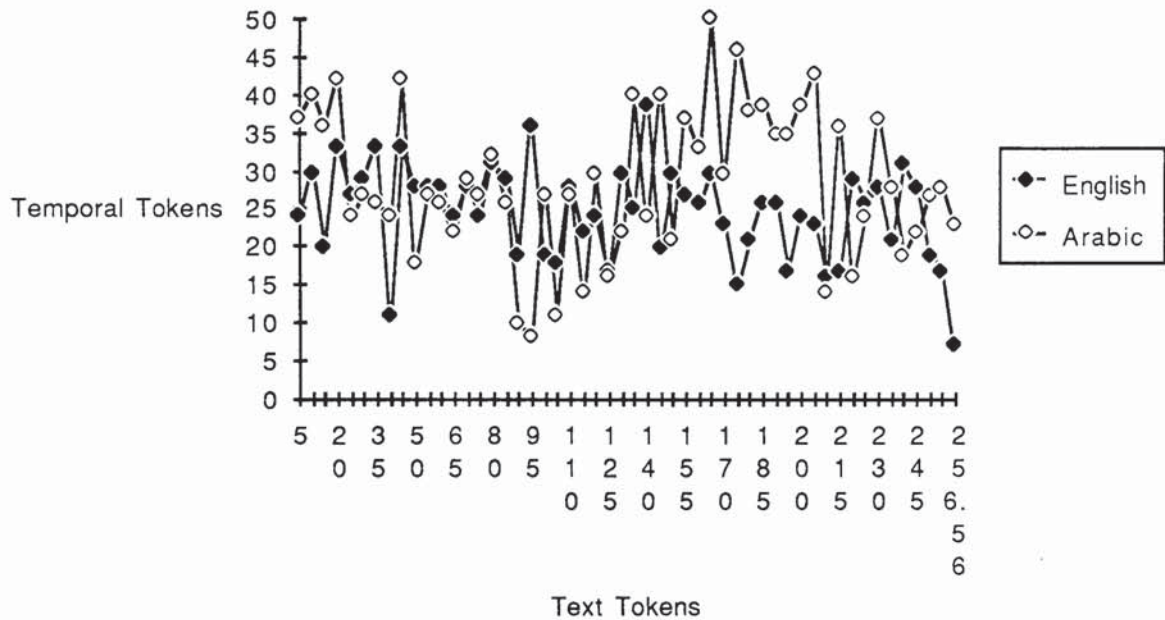
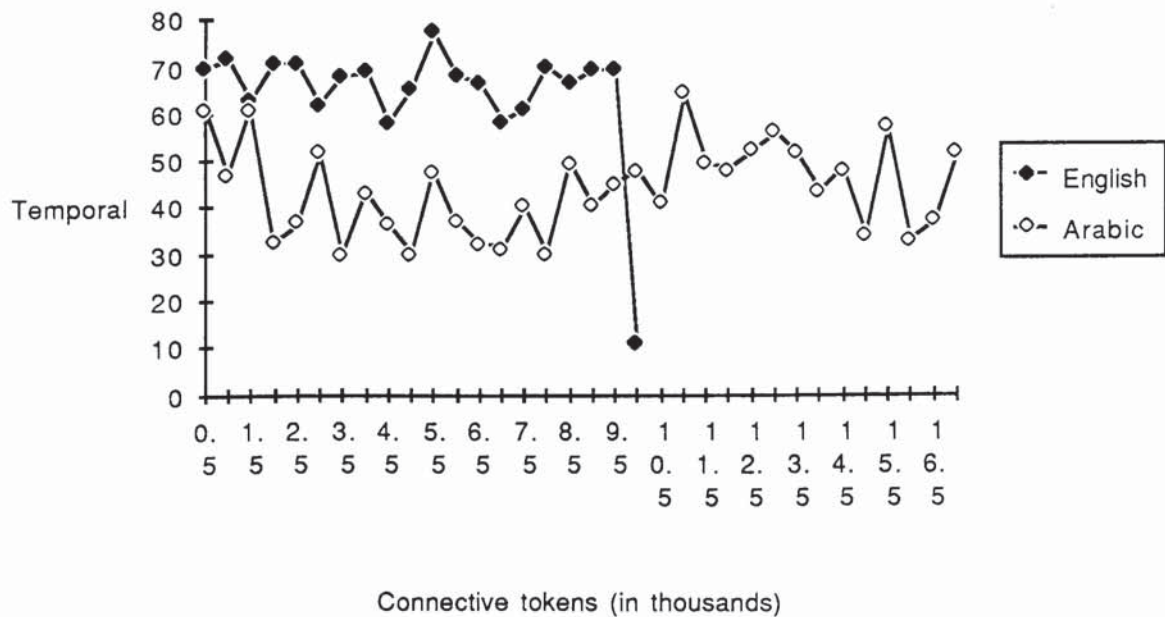


Fig. 9.44 Comparison of the Local Growth of Temporal Connectives



of extrapolation. A million word English corpus is expected to contain 5,119 temporal tokens, whereas an Arabic corpus of a similar size is expected to include a smaller number: 4,894.

ii. Calculation of growth of types shows faster saturation in English compared to Arabic. For instance, at interval 5 (25,000 words) growth achieves 50% of types in English and 38% in Arabic. At interval 15 (75,000 words), growth achieves 75% of types in English and 56% in Arabic. And at interval 28 (140,000 words) growth of types reaches 90% in English and 73% in Arabic. A million word English corpus is expected to include 74 types (increase of 54%), while an Arabic corpus of a similar length is expected to contain 110 types (an increase of 77%).

b. Local Growth

Calculation of Tuldava's index of richness of temporal connectives within text intervals each comprising 500 connective tokens gives a higher index in English than Arabic (1.0703 in English compared to 1.02 in Arabic). Growth rate index is only slightly higher in English than Arabic (-1.0842 vs. -1.077). However, since temporal tokens achieve a significantly higher relative frequency within English, results of extrapolation show higher figures than in Arabic. For instance, a corpus that contains 100,000 connective tokens is expected to have 12,912 temporal tokens, whereas an Arabic corpus containing the same number of connective tokens is expected to include 7,160.

4. Distribution of Temporal Categories

Comparison of the distribution of temporal categories in both corpora are displayed in Figures (9.45-46). Each category is considered separately below:

a. Temporal sequencing

Connectives of temporal sequencing in Arabic constitute a larger category in terms of tokens than their counterparts in English. They comprise 780 tokens, which represents 52% of temporal token mass. In English the category comprises 356 tokens, which correspond to 28% of temporal tokens.

In terms of types, connectives of temporal sequence contain 20 types in English, of which 2 are hapaxes. Their counterparts in Arabic comprise 18 types, of which 6 are hapaxes. Accordingly the connective set in English manifests higher concentration but lower variegation than Arabic.

The category differs across the two languages in the behaviour of the most frequent connectives. In Arabic, 3 connectives represent 72% of token mass. These are "wa", "ba<da" and "tamma". In English, a similar share is achieved by the combined frequency of 5 connectives: "and", "before", "then", "after" and "eventually".

b. Simultaneity

This is a larger category in English both in terms of absolute and relative frequencies and in the number of types. It constitutes 187 tokens in English representing approximately 15% of total

Fig. 9.45 Comparison of the Distribution of the Categories of Temporal Connectives in the two Corpora

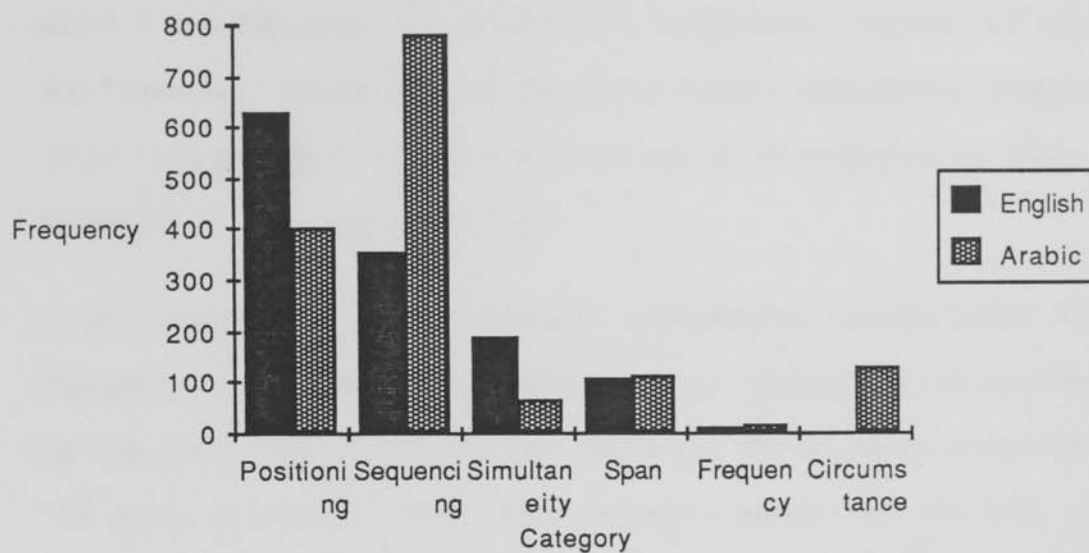
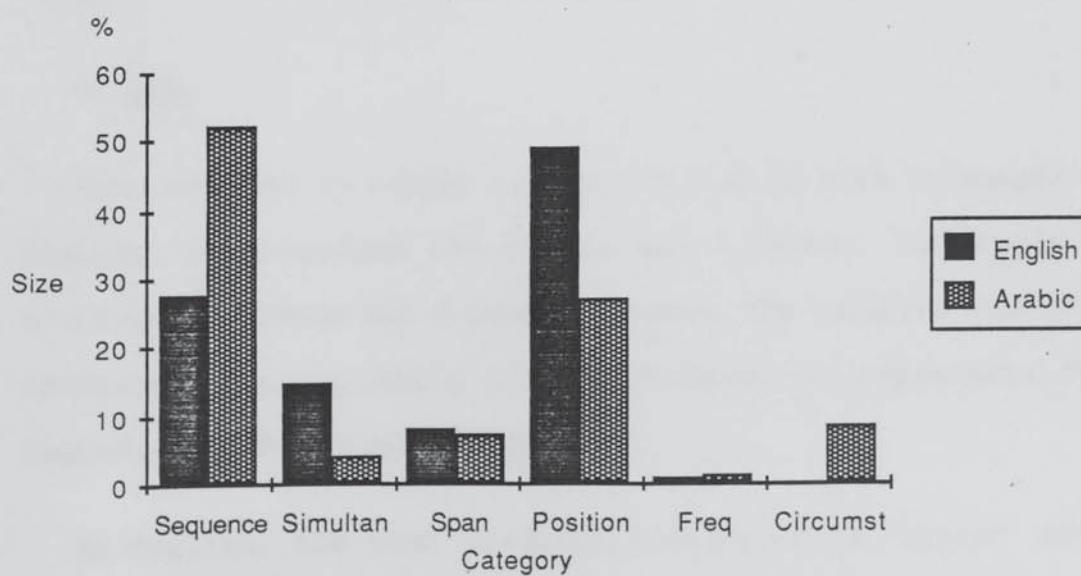


Fig. 9.46 Comparison of the Distribution of Shares of Temporal Categories



temporals, while in Arabic it comprises 62 tokens that stand for 4% of temporals.

In terms of types, the category contains 9 types in English, of which 2 are hapaxes. In Arabic, it comprises 7 types, of which 2 are hapaxes. Hence the category in Arabic manifests higher TTR (0.11 compared to 0.05), concentration (0.08 compared to 0.04), and variegation (29 compared to 22).

In Arabic the most frequent connective constitutes 73% of connective tokens of this category. This connective is represented by the core "waqti" and has the variants: "fī al-waqti nafsi-hi" and "fī nafsi al-waqti". This high frequency means that the rest of the types have lower frequency each. In contrast, English makes use of 4 connectives whose combined frequency represents 90% of tokens. These are "as" (38%), "while" (28%), connectives with "time" as a core: "at the same time" and "all this time" (12%), and "meanwhile" (12%).

b. Span

This category is nearly similar in size in both languages. In English, it comprises 104 tokens and 6 types. In Arabic, it contains 110 tokens and 6 types. However, the relative size of the category differs slightly in both corpora: it represents 8% of English, and 7% of Arabic, temporals.

In English, the most frequent connective is "since" (40% of tokens in this category) followed closely by "until" (38.5%). This indicates a similar proportion between connectives expressing

"backward" or "past" span and those expressing "forward" or "future" span. In Arabic, however, the top two connectives are "ḥattā" (50% and "'ilā 'an" (24.5%), both expressing "future" span. This means that Arabic exploits other means than connectives in the expression of past span.

c. Positioning

This is a larger category in English in terms of the absolute and relative frequencies of tokens. It comprises 626 tokens that represent 49% of English temporals. In Arabic, it comprises 400 tokens that stand for 27% of temporals.

In terms of types, this category includes a smaller number of types than Arabic: it comprises 22 types in English, of which 5 are hapaxes, and 37 in Arabic, of which 13 are hapaxes. Accordingly, the category in Arabic manifests higher TTR (an index of richness and extensity of use of types, calculated as 0.09 vs. 0.035), higher concentration (0.06 vs. 0.03) and variegation (35 vs. 23).

The category in English is dominated by the connective "when" (66% of tokens in the category) and "then" (10%). In contrast, this large share of tokens is represented by 8 connectives in Arabic: "<indamā", "hīna", "ḥīnamā", "'idā" (as a temporal connective), the core "waqti" in the compound connectives "fī ḥādā [ḍālika] al-waqti", "<inda'idin", "yawma", and "'ānaḍāka".

e. Frequency

This is a small category in both corpora. In English, there are 11 tokens and 2 types: "whenever" (9 tokens, equal to 82%) and

"every time" (2 tokens, 18%). In Arabic, the category comprises 17 tokens and 4 types: "kullamā" (simple connective, 13 tokens equal to 76%), "kullamā ... kullamā" (2 tokens) and "mā dāma" and "mā", both hapaxes.

f. Circumstance

This is signalled in Arabic by the connective "wa", which has a frequency of 125, representing 8% of temporals. A similar counterpart has not been observed in English.

9.2.7.2 Textual Variations of Temporal Connectives

9.2.7.2.1 Temporal Sequence

A number of textual variations have been observed in the way English and Arabic connectives signal the temporal relation of sequence or succession. The most noticeable differences can be classified in four types: conditions of sufficiency and necessity, gradation of temporal ordering, multiple occurrence, and fuzziness in signalling. These are treated with some detail below.

1. Conditions of sufficiency and necessity

In Arabic, condition of sufficiency for the provision of explicit signalling is satisfied when there is a temporal ordering of two knowledge blocks in the text world, however vague or remote it may be. As with signalling the additive/appending relation in Arabic, the condition of sufficiency, once it is met, immediately creates a necessary pressure to utilise connectives for explicit signalling. In English, however, temporal succession needs to be well-defined in

order to satisfy the condition of sufficiency. Moreover, the fact that two knowledge blocks, be they facts, events, etc., are temporally ordered does not impose explicit signalling of the relation. The choice is a rhetorical one and relies on the extent that connective utilisation serves principles of textual design: efficiency, effectiveness or appropriateness. This factor explains the stronger tendency for explicit signalling of temporal sequencing in Arabic compared to English.

2. Fuzziness of signalling

The strong tendency discussed in (1) above for signalling temporal sequencing via connectives in Arabic relative to English creates some fuzziness in the semantic scope of the relation. This fuzziness is of two related types; each is associated with some connectives. One type is attributed to reduction of functional distinctness while the other results from complexity of signalling. These are explained below:

a. The relation is "played down", i.e. its definition is minimised, via the use of "wa" in Arabic. Since this connective is essentially additive, its use reduces the explicitness of succession and deemphasises the relation it encodes. There are numerous examples in the corpus that exemplify this tendency. For instance in this excerpt:

- [9.17] wa qad <aqad-ā 'ijtimā'an muštarakan fī Al-qāhirati fī 'uktūbar 1977 taḥta šī<āri al-taḍāmuni min 'ajli taqaddumi wa rafāhiyyati 'abnā'i wādī al-nīli.
 wa jā'a 'ijtimā'u-humā al-tānī fī al-xurṭūmi fī yanāyir 1979 taḥta nafsi al-šī<āri.

[They [the Egyptian and Sudanese Parliaments] held a session in Cairo in October 1977 under the slogan "Solidarity for the progress and welfare of the peoples of the Nile Valley".

And their second session was held in Khartoum in January 1979 under the same slogan ...]

(Hr, 28/5/1983, X61, 11100-6)

the connective "wa" establishes a relation of temporal succession conflated with additivity. This conflation relatively obscures the distinctness of the temporal signal. It should be noted that a similar effect is created in English when "and" is used in a temporal sense. However, as mentioned in (1) above and as the frequency distribution of connectives of this category confirms, the tendency to resort to fuzziness is stronger in Arabic than in English.

b. Related to (a) above is the tendency of connectives other than "wa" in Arabic to conflate one or more semantic functions with temporal succession. But while the conflation of functions in (a) leads to relative indistinctness, it contributes in this case to richness of signalling. In other words, the complex nature of the relation dictates multiple functionality which in turn tends to create fuzziness in the expression of temporal succession. Connectives that may at times be associated with this type of fuzziness are "fa" and "tamma". For instance in the following excerpt, temporal sequencing which "tamma" signals is conflated with adversativity:

[9.18] wa taraddada fī al-'awsāṭi al-diblūmāssiyati al-hādi'ati 'anna al-'insihāba al-'isrā'iliyya min Lubnān sa-yatimmu qabla nihāyati <āmi 1982. tamma mā labiṭa 'an ta'axxara al-maw'idu šahrān 'itṭra šahrin.

[It was repeated in the moderate diplomatic

circles that the Israeli withdrawal from the Lebanon will be completed before the end of 1982. (However, then) this date was delayed one month after another.]

(Hr, 25/5/83, X58, 10863-7)

Again, fuzziness of this type does occur in English, but the tendency is stronger in Arabic. The connective "fa" itself is multi-functional and may sometimes conflate a number of functions to indicate one complex textual relation.

3. Gradation of succession

In Arabic, the connectives "fa" and "tamma", despite their occasional complexity of functioning, which we discussed in (2) above, normally indicate a distinction of time lapse between two events, acts, etc. The connective "fa" points to immediate succession, i.e. it indicates that the time lapse is very short. This has the effect of imposing a rapid gradation of temporal succession. For example, each of the following two excerpts contains two events that happen in a rapid sequence:

- [9.19] ... <amūdun ma<daniyyun taḍa<u-hu fī kūbi al-mā'i
fa yataḥawwalu 'ilā tiryāqin šāfin
[... a metal bar that, once you immerse in a
glass of water, changes it into a cure.]
(Hr, 24/1/83, X28, 5855-6)

- [9.20] wa la<alla-ka tas'alu nafsa-ka: hal ṣaḥaw-ta fī
yawmin mā fa wajad-ta nafsa-ka šā<iran, hākaḍā,
bi-quwwatin siḥriyyatin ...?
[You may ask yourself: did you wake up one
morning and then, with some magic force, found
yourself a poet ...?]
(J, 27/5/83, X101, 19253-5)

The connective may be utilised a number of times, thus ordering blocks of knowledge in a temporal series. The organisational force

of this use implies some classificatory arrangement across the temporal dimension. Examine this excerpt:

- [9.21] ... ša'at mušādafatun 'an yamurra dālika al-rajulu min hunāka fa yašhada al-mar'ata al-qiddisata fī miḥnati-hā, fa yakādu yuṣābu bi-al-'igmā'i faza<an mimmā yarā.
[Fate had it that that man would pass in that direction, and see the holy woman in her distress; he was about to faint with horror at what he saw.] (Hr, 24/4/83, X53, 9664-7)

Related to this function is the use of "fa" to indicate roles of speakers in a reported conversation. After the text producer reports or states what the first speaker has said, he uses "fa" to connect it to the second speaker's response. By way of exemplification, this excerpt describes a conversation between the American Foreign Secretary Mr. Shultz and his British counterpart Mr. Pym.

- [9.22] wa qāla šultz li-zamīli-hi al-barīṭāniyyi: - 'inna širā<a al-šarqi al-'awsaṭi huwa fī al-ḥaqīqati širā<un bayna al-dawlatayni al-'a<zami ...
fa 'ajāba al-wazīru al-barīṭāniyyu bi-'anna hādā al-waṣfa gayaru ṣaḥiḥin ...
fa radda al-wazīru al-'amrikiyyu bi-'anna al-halla al-maṭlūba lan yakūna maqḥūlan min al-furaqā'i al-ma<niyyīna 'illā 'idā wāfaqat <alay-hi al-dawlatāni al-'a<zamu ...
[Mr. Shultz said to his British counterpart, "The Middle East conflict is in reality a conflict between the two superpowers ...".
(Then) The British Minister replied that this description was not true ...
(Then) The American Foreign Secretary answered that the required settlement would not be accepted by the concerned parties unless the two superpowers approved it first ...]
(Nb, 15/7/83, X22, 4575-95)

The text writer starts by quoting Mr. Shultz's statement. Mr. Pym's reply is marked sequentially with "fa", meaning roughly "and then",

and thus it is connected to the first (i.e. Mr. Shultz's) quotation. The response to Mr. Pym's comments is also marked sequentially. The result is a textual series the components of which are sequentially marked, ordered and chained to each other.

Unlike "fa", the connective "tūmma" indicates a longer time lapse in succession and, therefore, lacks its characteristic immediacy of temporal ordering. In general "tūmma" implies a pause and an interval separating ordered events, acts or facts. For instance in this excerpt

- [9.23] wa hākaḏā mātat mubādaratu Rīgin ... tūmma mātat
'ittifāqiyyatu Lubnān ...
[Thus Reagan's initiative perished ... Later the
Lebanon treaty perished ...]
(Nb, 15/7/83, X22, 4312-5)

there is a time interval between the two events. The succession is reinforced further through the use of parallelistic structures and repetition of the verb "mātat" [perished, lost life].

In order to stress either of the two combined functions that "tūmma" signifies : "pause", and "interval", text producers support it with appropriate lexical items or other sequential connectives. In the following excerpt

- [9.24] wa šamata šadiq-ī. tūmma 'istaṭrada qā'ilan ...
[My companion paused. Then he went on saying
...]
(Nb, 5/4/83, X8, 991)

the use of the verb "šamata" [paused] stresses the meaning of pause as expressed by "tūmma". In the following excerpt:

- [9.25] ... fī al-bidāyati kānat tajidu 'ihtimāman tumma 'aṣḥaḥat rūṭīnan min rūṭīni al-ḥayātī al-siyāsiyyati.
 [At first they [the Geneva talks] received some interest; then they turned into one of the regular routines of political life.]
 (Ar, 17/6/83, X211, 36344-6)

succession is reinforced in two ways: a) the meaning of interval is enhanced through the use of "fī al-bidāyati" [at first] (otherwise the adversative meaning associated with "tumma" in this particular context will be dominant), and b) the use of two parallelistic and highly balanced clauses.

Like "fa", the connective "tumma" can be involved in setting up a close temporal arrangement in the textual world whereby the repetition of "tumma" creates a chain of successive acts, events or facts. This is exemplified in [6.136] in Ch. 6 above.

Another example is the following excerpt:

- [9.26] ba<du-hum yaḍa<u xuṭṭatan tumma yu<addilu-hā qabla 'an yabda'a tanfīḍu-hā, tumma yamaṭṭu fī marāḥili tanfīḍi-hā ...
 [Some [planners] put a plan, then they modify it even before starting to implement it, then they unnecessarily prolong the stages of its implementation ...]
 (Nb, 21/5/83, X12, 2044-6)

The sequential gradation that "fa" and "tumma" signify is illustrated better in the following excerpt where a short interval between two events is signalled by "fa" while a longer one is indicated by "tumma":

- [9.27] wa kāna kitābu (Hasan Tawfiq Al-<adl) ... 'awallu kitābin <arabiyyin fī hādā al-<ilmi. tumma tatālat al-kutubu ba<da-hu. fa zahar kitābu (Jurjī Zaydān) ... fī sanati 1911, wa kitābu

(ʾahmad Ḥasan Al-zayyāt) alladī nušira baʿda kitābi Zaydān bi-ʿiddati šuhūrin.

[The first Arabic book in this field was by Hasan Tawfiq Al-Adl. Then similar contributions started to appear: a book by Jurji Zaydan appeared in 1911 and, a few months later, another one by Ahmad Hasan Al-Zayyat].

(J, 27/5/83, X102, 19438-44)

A similar distinction of connectives on the basis of gradation of succession is missing in English, although, we must note, there are other means of signalling it. The semantic scope and organisational force of the connective "tamma" may be approximated by "then" or "and then"; but there is no connective that can act as a counterpart to "fa". Since this connective plays a significant role in ordering the temporality of the events in the text world, and, more generally, in the rhetorical organisation of text segments, it follows that a missing counterpart results in a distinctive textual variation in both languages.

4. Multiple Occurrences

A strong tendency in Arabic, and this is verified and confirmed through two OCP runs, is the multiple occurrence of connectives of succession. This refers to the significant reliance of temporal organisation in the text on the explicit signalling via connectives of this category. They are used repetitively and in close proximity to an extent that is not tolerated in English. The concordances that we have generated reveal that in Arabic there are 187 instances of two connectives of temporal sequencing occurring within a distance of up to 20 words from each other. These instances comprise 36 combinations of connectives, as displayed in Table (9.3). In English this number is significantly smaller; there are

1st conn. in combination -----	2nd conn. in combination -----	Freq. ----
'axīran	fa	1
'axīran	wa	1
'awwalan	ba<da	1
'awwalan	tumma	2
bad'an	tumma	1
ba<da	'axīran	1
ba<da	ba<da	3
ba<da	fa	1
ba<da	qabla	1
ba<da	wa	9
fī al-bidāyati	tumma	1
tumma	'axīran	2
tumma	ba<da	4
tumma	tumma	6
tumma	fa	3
tumma	qabla	1
tumma	wa	12
fa	'awwalan	1
fa	ba<da	2
fa	tumma	1
fa	fa	7
fa	qabla	3
fa	wa	13
qabla	'axīran	1
qabla	fī al-bidāyati	1
qabla	tumma	3
qabla	fa	2
qabla	qabla	2
qabla	wa	3
wa	'axīran	1
wa	ba<da	9
wa	fī al-bidāyati	1
wa	tumma	8
wa	fa	14
wa	qabla	2
wa	wa	63
Total combinations: 36		
Total Frequency: 187		

Table 9.3 Frequency distribution of connectives of temporal sequencing occurring within a distance of 20 words in the Arabic corpus

only 14 combinations that have a total frequency of 31, as shown in Table (9.4).

1st conn. in combination -----	2nd conn. in combination -----	Freq. ----
after	finally	1
and	and	5
and	before	1
and	eventually	1
and	then	7
before	and	1
before	then	4
at the end	and	1
at the end	then	1
finally	and	1
first	then	2
later	eventually	1
later	in retrospect	1
then	and	4
Total combinations:	14	
Total Frequency:	31	

Table 9.4 Frequency distribution of connectives of temporal sequencing occurring within a distance of 20 words in the English corpus

9.2.7.2.2 Simultaneity

The quantitative behaviour of connectives of this category underlies two textual variations across the two languages. The first is related to the general tendency of utilisation while the second concerns the type of connective, and therefore the type of organisation, that is utilised for the expression of the relation.

1. Tendency in Utilisation

According to the frequency distribution of the category in both corpora, there is a stronger tendency in English a) to utilise connectives for signalling temporal simultaneity of events, courses of actions or states, and b) to signal both subfunctions of simultaneity (cf. 6.9.3 above): simultaneity "proper" and "abutting" simultaneity.

In the English corpus connectives that signal simultaneity proper organise events, courses of actions, etc. in 3 distinctive structures:

a. Span-Span Simultaneity: The connective signals simultaneity of two activities with a large measure of overlap between them, i.e. they are conceived of as starting and stopping at about the same time. The following excerpt exemplifies this type:

[9.28] ... the crowds sang ... as they waited to greet
ships returning from the Falklands.
(G, 17/1/83, X13, 1801-3)

Position of the connective does not interfere in the overlap of the two activities, though it may have some effect on linearisation, particularly the way focus is applied and where it is placed, the kind and amount of regression or progression required, and the type of grouping that the sequence will enter into.

b. Event-Span Simultaneity: In this type of organisational structure, a punctiliar event takes place during a span, i.e. within the temporal duration of another event, course of action or state. Lineally, the event and span configurations are organised in one of

two ways:

i) The span configuration may be set within the antecedent, which is usually introduced by the connective. Inspection of this type of organisation in the English corpus suggests that the span often serves as a background against which the event is highlighted. For example.

[9.29] One evening last summer I heard a tremendous rumpus emanating from a clump of trees in the cow's field. As I approached, a little bull came flying out of it, all four feet in the air.
(DTel, 18/4/83, X104, 17442-5)

ii) The span configuration, often still introduced by the connective, may be set within the subsequent. In this structure, the event receives prominence by being placed, and therefore processed, first.

[9.30] ... if she's lucky, [she can] get an hour or two off while they are processing the film.
(STi, 23/1/83, X233, 36701-2)

c. Event-Event Simultaneity: In this structure the connective establishes a relationship whereby two punctiliar events are presented as taking place at the same time. For example,

[9.31] ... he finally remarked to me as we parted after a fairly gloomy conversation ...
(O, 15/5/83, X66, 10828-30)

A variant structure that is related to (1) and (2) above is where the subsequent does not express a clearly punctiliar event nor a distinctly continuous span. Instead, it represents a course of

action or a state that develops, rather than happens, during a span that is expressed in the antecedent. This is exemplified in the following excerpts:

[9.32] As the country [Soviet Union] moves towards a new attack on economic difficulties and greater and overdue attention to the production of consumer goods, it finds itself faced by the likelihood of greater military demands.

(O, 15/5/83, X65, 10543-8)

[9.33] As the months go by, Britain should look a more attractive haven for international money than many other places ...

(STel, 12/6/83, X145, 24288-90)

In contrast, the Arabic corpus displays a lesser textual pressure for using connectives to organise text sequences on the basis of temporal simultaneity and coincidence. This is evident in the far lower frequency of tokens and types of connectives in this category and the share that this category represents in the total mass of Arabic connectives. The function is more often expressed in circumstantial constructions introduced by "wa" (see 6.9.6).

Further, a large percentage of the connectives signal relations of "abutting" simultaneity, while only a small number of connectives signify relations of simultaneity proper. These, unlike their counterparts in the English corpus, display fewer organisational structures, the main one being the event-span simultaneity and its variant structures. It should be noted that connectives such as "baynamā", "fimā" and "'id" are capable of creating the other structures, but their absence from the corpus points to a tendency that is peculiar to text organisation in Arabic, and therefore constitutes a distinctive variation.

2. Type of Operationality

Observation of connective types of this category and their frequency reveals a difference in their construction. In English the relation of simultaneity is more frequently signalled by the subordinators "as" and "while". Other means are available but to a far lesser extent. In Arabic, the signalling of the relation is dominated by the use of connectives with anaphoric element, particularly "fī nafsi al-waqtī" [at the same time] or its variants "fī al-waqtī nafsi-hi (dāti-hi)", or "fī dālika (hādā) al-waqtī" [at that (this) time]. (But refer to the discussion of circumstantial "wa" in 6.9.6; see also 9.2.7.2.5 below)

This is a significant difference as it influences the manner of text linearisation in one language relative to the other. Some of the differential manifestations concern salience while others are related to the various activities involved in textual sequencing (cf. Ch. 3 for a discussion of the principles of text linearisation).

a. The Imposition of Salience: The use of subordination distributes the flow of control in such a way that the control centre, which carries the textual core (cf. 3.6.3.3), is anchored at the main clause, thus placing its content in focus. The dependent clause would routinely be peripheral. This explains why in establishing temporal overlap, the content of the dependent clause that is introduced by "as" or "while" acts as a temporal background for the content of the main clause. On the other hand, the use of a

connective with an anaphoric structure distributes the control centres according to their ideational plan in the textual world, and therefore the temporal overlap of background/foreground events is not necessarily relevant.

b. Activities of Sequencing: A connective with an anaphoric structure triggers off activities dominated by regression. The activities are normally controlled by the exigencies of processing, thus determining, for instance, how much text can be spanned or reactivated by the connective. In contrast, the subordinators "as" and "while" invoke either activities of progression if the dependent clause is activated first, or activities of regression if the dependent clause is delayed till after the main one.

The variations discussed above indicate a stronger tendency towards versatility in the expression of temporal simultaneity in the English compared to the Arabic corpus. Organisation is more variable, depending on whether a temporal orientation point is stated first and then the main prominent configuration, or vice versa.

9.2.7.2.3 Temporal Positioning

The main interlingual variations observed in the textual behaviour of connectives of temporal positioning can be grouped into two types. One concerns the nature of the connectives while the other is related to certain constraints of positioning.

1. Nature of Connectives

The following two observations on the nature of connectives of

this category are based on inspection of the quantitative profile and relevant concordances.

a. There are two types of connectives: subordinators such as "when" and "<indamā", and time relators with a referential element, e.g. "in those days" and "fī dālika al-waqtī" [at that time]. The cumulative frequency of each type in relation to the total number of connectives within this category is similar in both corpora: 73% subordinators and 27% time relators in English; 77% subordinators and 23% time relators in Arabic. However, in English the number of observed subordinators is smaller than in Arabic: 8 vs. 16. Temporal positioning in English is heavily signalled by the subordinator "when". A similar function in Arabic is signalled by a number of connectives: "<indamā", "hīna", "hīnamā", "yawma" and, to some extent, temporal "'idā" and "ba<damā".

2. Related to the previous point is the distinction between intensity as opposed to extensity of use of time relators (with referential elements) in the two corpora. The tendency in English is towards intensity of utilisation. That is, there are fewer cores that can be utilised for constructing a wider variety of connectives by varying the referential element. Inspection of the concordances reveals 9 cores that can be used for producing 28 different forms of connectives. These are shown in Table (9.5) below.

In Arabic, there is a tendency towards extensity of use. This means that, in contrast to English, there are more cores to be utilised for constructing connectives but the possibilities for utilising individual cores are severely restricted. If we exclude

1. circumstances	in such circumstances, in the circumstances in these circumstances.
2. day	in those days, on the same day.
3. event	in the event, in such an event, in that event.
4. moment	at such a moment, at that moment, at the moment, at those moments.
5. occasion	on this occasion, on that occasion, on the same occasion, on another occasion, on these occasions.
6. point	at just that point, at this point.
7. stage	at this stage, at that stage.
8. then	then, by then.
9. time	at the time, at that time, at a time, this time, at one time.

Table (9.5) Connective cores used for constructing time
relators with referential elements
(as observed in the English corpus).

1. 'id	'id dāka [at that time, then]
2. bad'	fī bad' [at the start]
3. bidāyat	min al-bidāyati [from the (that) start]
4. ḥālat	fī ḥādihi al-ḥālātī
5. ḥīn	fī ḥīni-hi [at that time]
6. zamān	fī dālika al-zamān [at that time]
7. ẓarf	fī ḥāḍā al-ẓarf [at this circumstance, time].
8. <inda	<inda-hā [at that moment, then]
9. fatrat	dāxila ḥādihi al-fatratī [within this period] fī ḥādihi al-fatratī [in this period]
10. marḥalat	fī ḥādihi al-marḥalātī [at this stage] min ḥādihi al-marḥalātī [from this stage]
11. marrat	ḥādihi al-marratā [this time], fī al-marratī [at (the) time].
12. laḥẓati	fī ḥādihi al-laḥẓatī [at this moment]
13. wāqi<	fī ḥāḍā al-wāqi<ī [in this event]
14. waqt	fī dālika al-waqtī [at that time]
15. yawm	fī al-yawmi nafsi-hi [on the same day].

Table (9.6) Connective cores used for constructing
time relators with referential elements
(as observed in the Arabic corpus).

morphologically composite connectives such as "'ānaḍāka", "ḥīna'idin", "<inda'idin" [at that time], "sa<ata'idin" [at that hour or time], "yawma'idin" [on that day] and the connective "hunā" used temporally, we then have a set of 15 cores that are used for producing 19 different forms of connectives as shown in Table (9.6) above.

3. Another observation concerns the absence of an Arabic equivalent to the connectives "as soon as" and "once". The functional scope of each of these two connectives combines precision of temporal positioning with rapid temporal succession of two events. The equivalent subordinator in Arabic, "ḥālamā", does not occur in the corpus. The nearest equivalent in the corpus is "laḥẓata" [the moment that], which occurs only once. A study of the connectives shows that this relation is more commonly signalled by "<indamā" [when], which, though it retains the combination of positioning and succession, lacks the subtle distinction of temporal rapidness and precise positioning.

2. Constraints on Positioning

These apply to the position of one type of temporal connectives, the subordinator. Survey of their behaviour in the Arabic corpus confirms that when they occur in initial position (and therefore the subordinate clause precedes the main one), the connective, while capable of maintaining interclausal linkage, creates a discontinuity in relation to the preceding text sequence, particularly the preceding sentence, and thus a textual gap emerges. This gap disturbs the stability of that part of the text to the extent that

processing activities are interrupted or brought temporarily to a halt. To restart the system, the processor has to perform extraneous operations of regression and progression to resolve and compensate for the missing link. This may unnecessarily overload the system, unless, of course, the textual disturbance is intentionally created for a certain rhetorical effect. To avoid the disturbance, the gap is immediately bridged with another connective, usually additive "wa", or, occasionally, "fa". Discontinuity does not occur when the subordinate clause is in medial or final positions. A similar constraint has not been noticed on the position of English equivalent subordinators.

9.2.7.2.4 Span

Survey of the behaviour of connectives in this category in the two corpora suggests two variations. One concerns the semantic scope of the connectives of future span and is related to their subtleties of signalling while the other considers positioning of connectives. These are discussed below.

1. Subtleties of Signalling

In English, the connective "until" operates on a textual continuum where at one end it signals a purely temporal meaning, and at the other it signifies a combination of temporality and causality. For instance, in excerpt [9.34] "until" is temporal, while in [9.35] it combines temporality with a shade of causality:

- [9.34] Take the large number of children trapped in care. No one was aware of this problem until Jane Rowe stumbled over it 10 years ago.
(G, 14/12/82, X8, 1133-5)

- [9.37] wa 'axaḡa hāḡā al-<adadu yakburu yawman ba<da yawmin. tumma 'imtadda ḡattā šamala ba<ḡa 'a<ḡā'i al-barlamāni.
[This number [of people unlawfully prosecuted in Iran] started to grow bigger. Then it extended until [to the extent where] it included some Members of Parliament.]
(J, 23/1/83, X80, 15690-2)

2. Positioning

There is a strong tendency in Arabic, as opposed to English, to support connectives of temporal span, particularly subordinators, when they occur in initial position in a text sequence. The connective is preceded by "wa" to reinforce its cohesive role and to relate it to the previous sequence. Connective support is vital in cases where the temporal connective starts to create discontinuity through failure to relate the next sequence with the current one. This problem is present to a variable degree of intensity in the way other types of temporal connectives, particularly those of positioning, function (see 9.2.7.2.3 above).

9.2.7.2.5 Circumstance

We have stated earlier (cf. 6.9.6) that the complexity of the textual features of this relation makes it peculiar to Arabic. The connective "wa", which signals temporal circumstance, may, in some instances, have a close semantic equivalence to "as", "while" or "inasmuch as". But often the equivalence is only partial, as inspection of the occurrences of circumstantial "wa" demonstrates. Two features are relevant: the functional complexity of the relation and some characteristic structural properties.

1. Functional Complexity

In signalling this relation, the connective establishes a complex semantic association between two statements. The association is basically temporal: the circumstantial statement has the primary function of describing a situation that is considered as an attendant circumstance to the event, course of action or state expressed in the main statement. There exists, therefore, a coincidence of occurrence between the content of the two statements in the textual world. This is, however, conflated with other functions, depending on the nature of the semantic relationship between the main event or state and its related circumstance. This relation is highly variable: it may be adversative, explanatory, orientative, or pure temporal (refer to the discussion and examples in 6.9.6.2). A similar English counterpart to the Arabic circumstantial "wa" has not been observed in the corpus.

2. Positioning

A study of the positioning of the circumstantial statement manifests three positions in relation to the main statement.

1. Final: The circumstantial statement follows the main one. In this position, the circumstantial statement may be one of two types:

a. directly related: There is a strong association between the content of the main statement and its circumstance. Generally, the content of the circumstance is either explanatory, descriptive, pure temporal or, to a lesser degree, orientative. Processing is

facilitated by a certain degree of predictability. The following is an illustrative excerpt:

- [9.38] *ṣawtu al-ṣabābi yuṣjīn-ī wa huwa yas'alu-nī <an al-ḥurriyyati wa ḥuqūqi al-'insāni.*
[The voice of the youths enchants me as (inasmuch as) they enquire about liberty and human rights]
(Sh, 29/11/82, X105, 20193-4)

b. detached: The association is weakened by the unexpectedness of the circumstance. In this case the circumstance is mainly in adversative relation to the content of the main statement. Processing requires activities of regression to update the conceptual knowledge in the text world. The detachment is sometimes orthographically marked by separating the main statement from its circumstance with a comma. The following excerpt is an example:

- [9.39] *fa kayfa ... lā yaqtani<u niṣāmu xumainī bi-dālika, wa huwa waḥda-hu alladī yujarribu layla nahāra talaqqiya al-ḍarabāti al-<irāqiyyati al-bāsilati ...*
[How could Khomeini's regime remain unconvinced when it experiences the brave and heavy Iraqi blows day and night?]
(J, 12/3/83, X88, 17060-3)

2. Medial: The circumstantial clause is often embedded within the main clause. In other words the circumstance is inserted as a parenthetical piece of information on the content of the main statement. This is sometimes orthographically marked by separating the inserted (nested) circumstantial statement from the main one by commas. The following is an example:

- [9.40] *... ma alladī mana<a al-sūfīta, wa hiya furṣatu-hum al-ḍahabiyyatu, min 'an yatadaxxal-ū bi-hazmin li-'īqāfi ḥarbi al-xalīji?*
[... what prevented the Russians, when that was

their golden opportunity, to have a firm
intervention to stop the Gulf war?]
(Ar, 18/4/83, X194, 33410-2)

3. Initial: This is a rare position. In fact, Cantarino (1975 Vol.III p.266), in discussing the subordinate function of the circumstance, states that "it is impossible for the subordinate to precede the main clause". This is, however, not accurate and the statement needs some amendment. We posit, in the light of the evidence we have from the corpus, that it is possible, but indeed rare, for a circumstantial to precede the main statement. The corpus contains one occurrence:

[9.41] wa 'anta taqra'u hādihi al-suṭūra, yaltaqī al'ān
fī al-<āsimati al-<irāqiyyati Bagdāda safwatun
min al-muslimīna wa kibāri al-mufakkirīna ...
[And as you read these lines, the elite of
Islamic scholars and a number of prominent
thinkers convene in the Iraqi capital, Baghdad..]
(Hr, 15/4/83, X49, 9039-41)

In this position, the connective "wa" combines an additive/continuative meaning with a temporal one. Note that while "wa" is here considered a subordinator it has a strong implication of coordination.

3. Multiple Embedding

Depending on the textual requirement of the main statement, it is possible sometimes to embed a number of circumstantial statements within the main one or within each other. However, it must be noted that processing will be overloaded, and, unless there is a strong motivation to utilise multiple embedding, it is not a frequent property. An example of multiple embedding is given in this excerpt:

- [9.42] ... fazi<-tu wa 'anā 'aqra'u suḥufa-nā al-
 <arabiyyata hādā al-'usbū<i wa hiya tanqulu <an
 wikālāti al-'anbā'i 'axbāra lajnati al-taḥqīqi
 al-'isrā'iliyyati ...
 [I was horrified as I read the reports in this
 week's papers of the Israeli investigation
 committee...]
 (Ar, 22/11/82, X158, 29531-4)

4. Constraints of the Structural Pattern

One of the essential properties of the circumstantial clause that is introduced by "wa" is its constructional pattern. The clause is usually nominal: the subject, unless there is an inversion, occupies the first place in the clause (refer to the examples in this section and in 6.9.6.2). If a verbal clause is used it has to be introduced by the particle "qad", or it is changed into a nominal one by placing a personal pronoun in the first position, i.e. before the verb, to refer explicitly to the subject (example [9.42] above).

These characteristic features of the circumstantial statement distinguish it from possible comparable structures in English. Indeed an equivalent in English would require selecting an appropriate one from a variety of different textual/structural patterns. This variation creates a difference of motivation during the phases of text actualisation (cf. Ch. 3) particularly the final two phases: conceptual development and expression. In English, one needs to search for and consider more choices and pick the structure that organises, enriches and elaborates ideation in a way that the final intended product satisfies requirements of textual design (cf. Ch. 3). In contrast, the complex functionality, positional mobility

and the possibility of embedding of the circumstantial structure in Arabic reduces comparable activities and thus help processing to get a faster access to a planned goal.

9.2.8 Spatial Connectives

9.2.8.1 Quantitative Variations of Spatial Connectives

1. Frequency

a. This is a small category in both corpora. In English, it comprises 50 tokens, which represents 0.5% of connective tokens in the corpus. In Arabic, it comprises a set of 40 tokens, equal to 0.2% of connectives in the corpus. In other words, the category is smaller in Arabic, both in absolute and relative frequencies, than in English. Figure (9.48) displays a comparison of the distribution of spatial connectives in the corpora. Figure (9.49) plots the distribution of spatial connectives against connective tokens.

b. The category comprises a set of 6 types in English, of which 3 are hapaxes. In Arabic, the category contains 3 types, of which only one is a hapax. Accordingly, the category in English manifests higher TTR (0.12 vs. 0.075), concentration (0.06 vs. 0.05), exclusivity (0.06 vs. 0.025), variegation (50 vs. 33) and density (6 vs. 5.458). Index of gravity is 17 compared to 40 in Arabic. The types show higher rhythmicality and stereotypicality of use in Arabic than English, the indices being 18.5 vs. 14.67 and 19.5 vs. 15.67 respectively.

c. In Arabic, the category hinges on the use of the "ḥaytu", the

Fig. 9.48 Comparison of the Distribution of Spatial Connectives

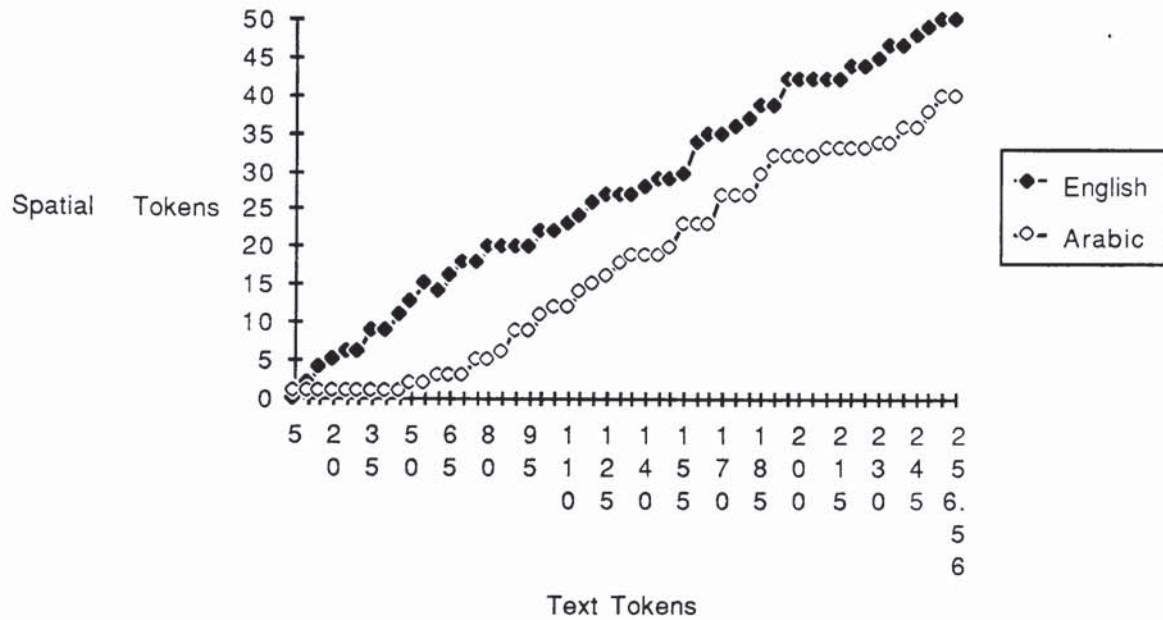
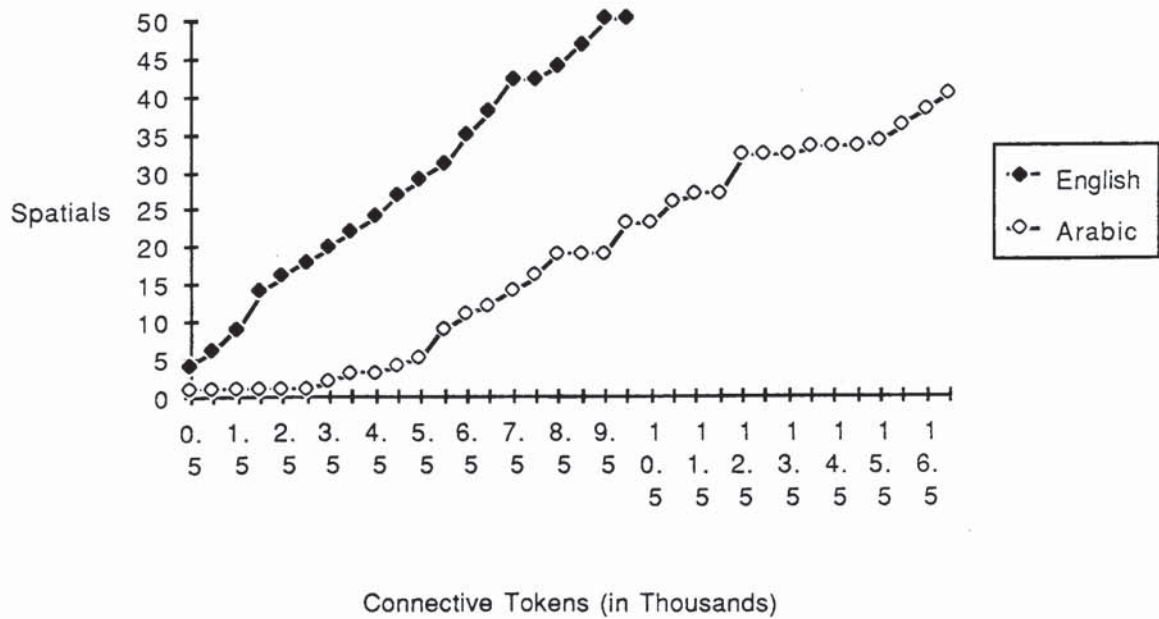


Fig. 9.49 Comparison of the Distribution of Spatial Connectives within Connective Tokens



frequency of which constitutes 93% of spatial tokens. In English, a similar relative frequency is achieved by the 3 non-hapax types: "where" (64%), "elsewhere" (16%) and "wherever" (14%).

2. Repetitiveness

In general, this category manifests comparatively higher repetitiveness in English than Arabic. This is evident in all the indices of repetitiveness that have been calculated. Spatial occurrence rate is 192 in English and 425 in Arabic. General repeat rate is (27×10^{-6}) in English and (5.5×10^{-6}) in Arabic. Spatial system repeat rate is (124×10^{-7}) in English and (48×10^{-7}) in Arabic. The average gap separating two occurrences of spatial connectives in English: 5,026 vs. 6,253. Type occurrence rate is 8 in English compared to 13 in Arabic.

3. Growth

Comparisons of global and local growth of spatial connectives across the two corpora are displayed in Figures (9.50-51) respectively.

a. Global Growth

i. In general growth of spatial tokens is relatively slower in Arabic than English. For instance, the first 9 intervals (45,000 words) contains one spatial token in Arabic compared to 9 in English. At corpus mid-point, English achieves 54% of token coverage, while Arabic achieves 40%. However, growth moves faster in the later intervals. According to this manner of growth, a million word English is expected to contain 183 spatial tokens in

Fig. 9.50 Comparison of "Global" Growth of Spatial Connectives

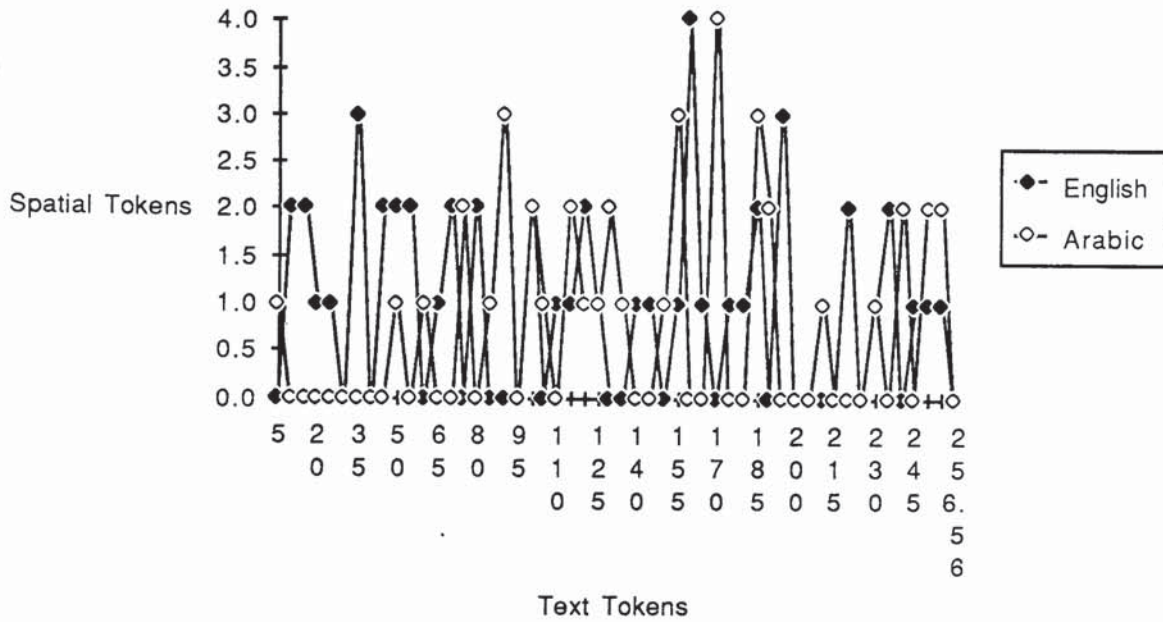
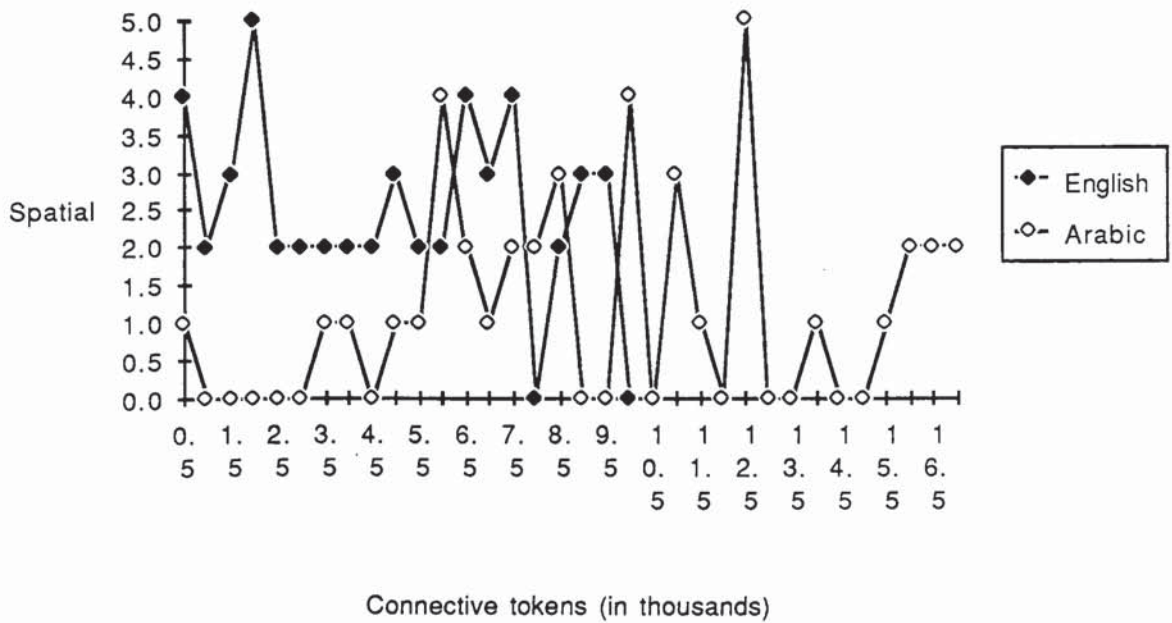


Fig. 9.51 Comparison of the Local Growth of Spatial Connectives



English and 202 in Arabic.

ii. Types show faster saturation in English, partly due to the manner of growth of tokens. For instance, the 67th type percentile appears in interval 3 (15,000 words) in English and interval 37 (185,000 words) in Arabic. Saturation occurs at interval 29 (145,000 words) in English and interval 38 (190,000) in Arabic.

b. Local Growth

Discrepancies in local growth are manifested in the first interval, where 4 spatial tokens appear in English and only one in Arabic. Growth ceases in Arabic until interval 7 (3,500 connective tokens) where another token occurs. At the same interval in English 20 tokens will have appeared. Using Tuldava's formula, the growth rate is -4.027 in English, higher than its Arabic counterpart, which is calculated as -13.006. Extrapolation to a corpus containing 100,000 connective tokens gives 407 spatial connectives in English and 370 in Arabic.

9.2.8.2 Textual Variations of Spatial Connectives

The interlingual variation in the behaviour of connectives can be summarised in two points. The first is related to their quantitative profile and its relation to their overall utilisation, while the second involves textual patterning.

1. Variation in Utilisation

Structurally, space relators are realised by a range of constructions of various sizes. What interests us here is the role

of some connectives in signalling a relation of space positioning between two propositions. This role is more operational in the English than Arabic corpus. For despite the fact that this category of connectives is the smallest in frequency in both corpora, the tendency to signal the relation is stronger in English. Evidence for this is the relatively larger share in connective mass that this category has in English compared to Arabic. Further, it has already been observed (cf. 6.10 above) that the distribution of connectives of this category within the Arabic corpus manifests a clear clustering in texts derived from Iraqi newspapers. A further investigation of this observation is referred to future work. It suffices to mention here that the relation is more evenly realised within the English than Arabic corpus.

2. Variation in Patterning

A closer inspection of the patterns of the two most frequent connectives "where" vs. "ḥaytu" manifests a degree of similarity. However, "where" has a pattern of connection that "haytu" has not been observed to have in the Arabic corpus. This pattern occurs twice in the following excerpt (cf. also, example 6.167 in Ch. 6):

- [9.43] A breakdown of the English Labour marginals shows ... that the better the Alliance did the higher was the swing to the Conservatives. Where the Alliance rise was less than average, the average two-partly swing was 3.9 per cent; where it was higher, the swing jumps to 6.8 per cent.
(STel, 12/6/83, X143, 23946-52)

It should be noted that a pattern similar to this one and having the same position of connective can be established with the use of

"ḥaytu" or "ḥaytumā" (probably preceded by additive "fa" or "wa", as a rendering of the above excerpt in Arabic would require). However, such a pattern is not present in the corpus.

9.2.9 Causal Connectives

9.2.9.1 Quantitative Variations of Causal Connectives

1. Frequency Distribution

a. This is a larger category in Arabic in terms of the number of tokens and types. It comprises 3,778 tokens that represent 22% of connective tokens, making it the second largest category in the corpus. In English, the category comprises 1,888 tokens that represent approximately 20% of connective tokens, making it the third largest category in the corpus. Figure (9.52) displays a comparison of the distribution of causal connectives in the two corpora. Figure (9.53) plots the distribution of causal connectives against connective tokens.

b. In Arabic the category comprises 76 types, of which 15 are hapaxes. In English, the category contains 45 types, of which 9 are hapaxes. These figures suggest slightly higher concentration in English (calculated as 0.019 compared to 0.016), higher exclusivity (0.005 vs. 0.004), and close variegation (20 compared to 19.7). Further, the category in Arabic shows higher rhythmicality and stereotypicality in the use of types (respectively 61.7 vs. 51.2 and 62.7 vs. 52.5 in English). Density is higher in Arabic, calculated as 11.9 compared to 10.9 in English, and gravity is lower 252 compared to 210 in English.

Fig. 9.52 Comparison of the Distribution of Causal Connectives

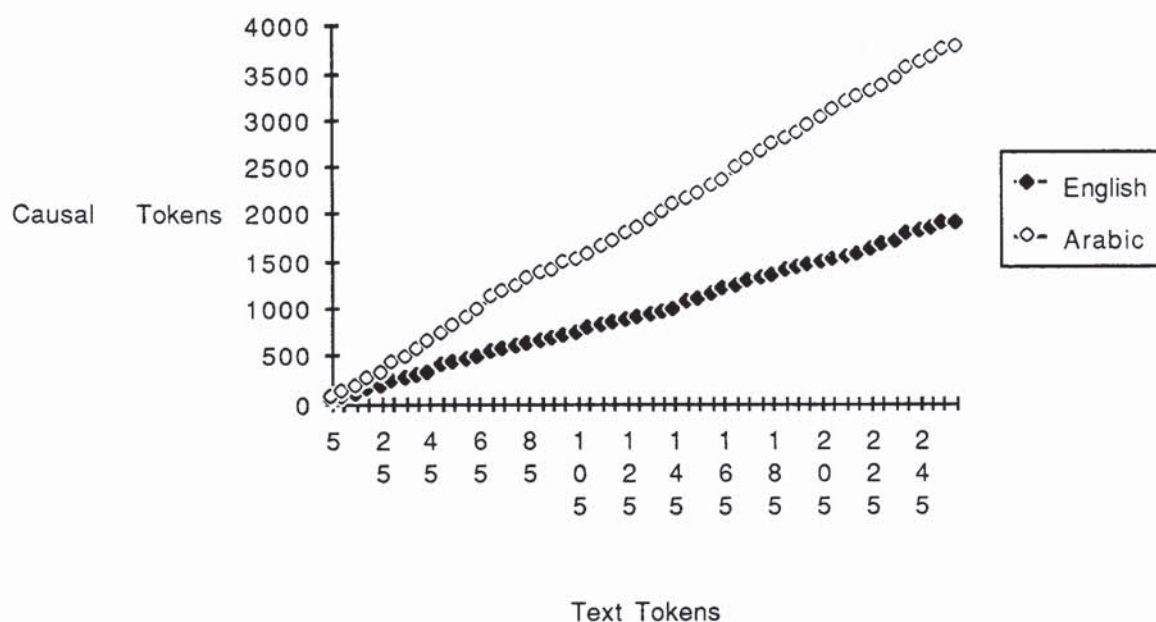
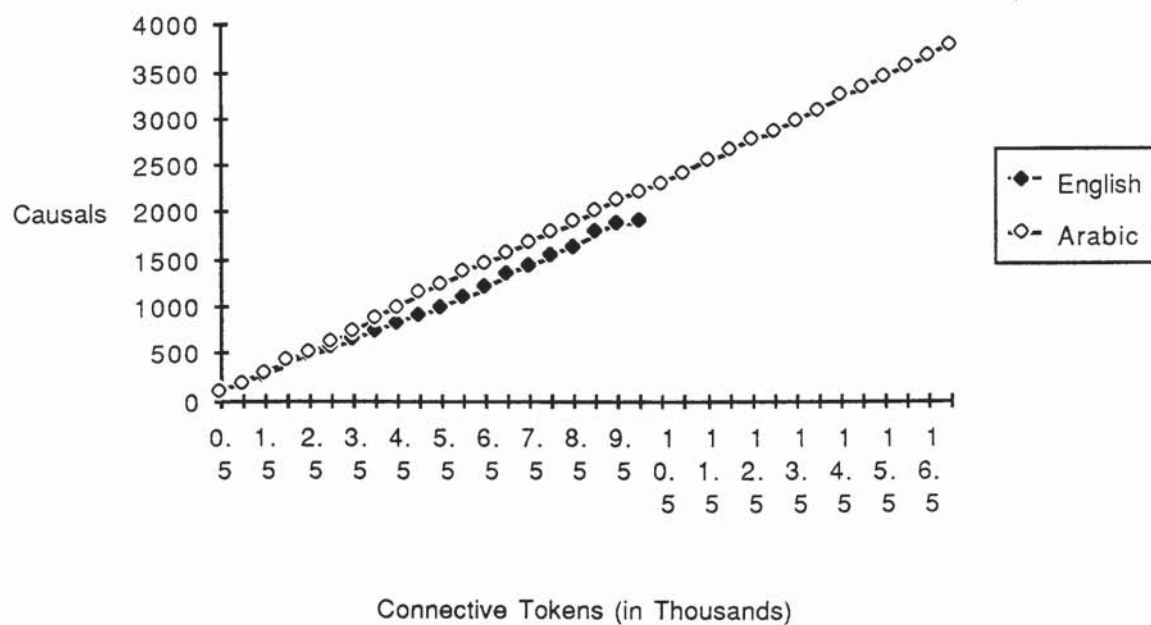


Fig. 9.53 Comparison of the Distribution of Causal Connectives within Connective Tokens



c. In English the category is dominated by the top ten types, the combined frequency of which constitute 90% of causal occurrences in the corpus. These are "if", "because", "and" (in a causal function), "so", "for", "then", "therefore", "since", "thus", and "as". This high frequency is represented in Arabic by the combined frequencies of 19 connectives (see App. 126 for details).

2. Repetitiveness

Causal connectives in both corpora show some similarity in certain patterns of repetitiveness and some difference in others. Causal occurrence rate and the general and system repeat rates are nearly similar. Causal repeat rate is 5 in English and 4.5 in Arabic; general repeat rate is 0.039 in English and 0.049 in Arabic; and system repeat rate is (66×10^{-4}) in English and (65×10^{-4}) in Arabic. On the other hand, type occurrence rate and the gap distribution are different. Type occurrence rate is higher in English: 42 compared to 50 in English. The average distance separating two causal connectives is lower in Arabic: 67 vs. 135.

3. Growth

Comparisons of global and local growth are plotted in Figures (9.54-55) respectively.

a. Global Growth

i. Growth of causal tokens in the English corpus manifests a certain extent of similarity to that in Arabic. Tuldava's growth rate is almost identical: -4.0756 in English and -4.0700 in Arabic. To investigate this further, we have looked into the intervals where

Fig. 9.54 Comparison of "Global" Growth of Causal Connectives

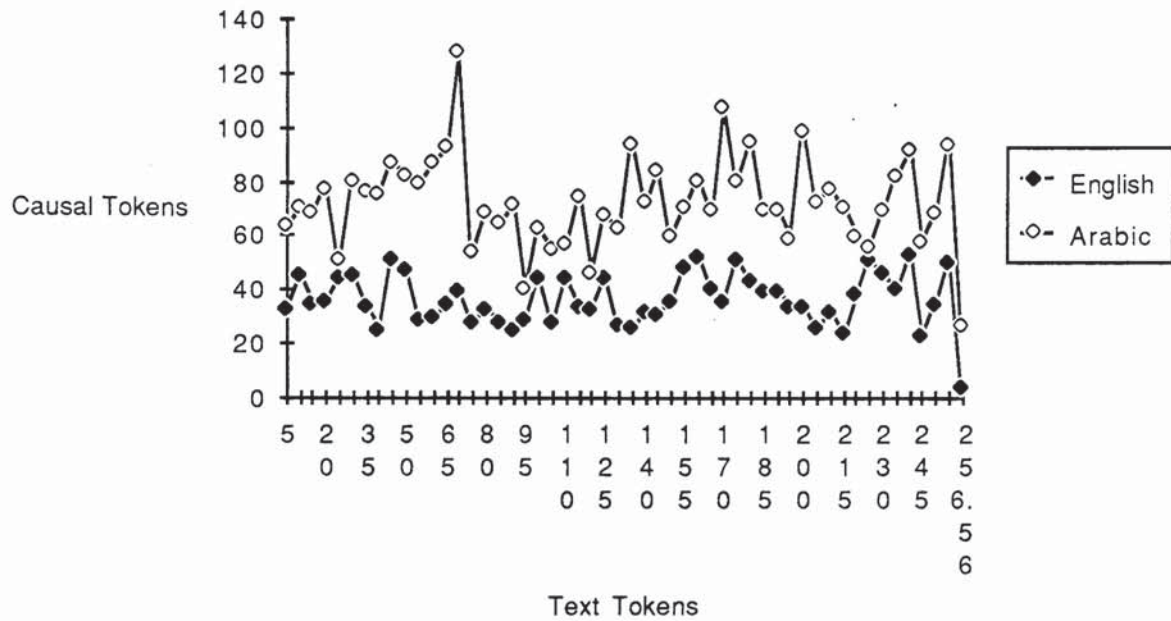
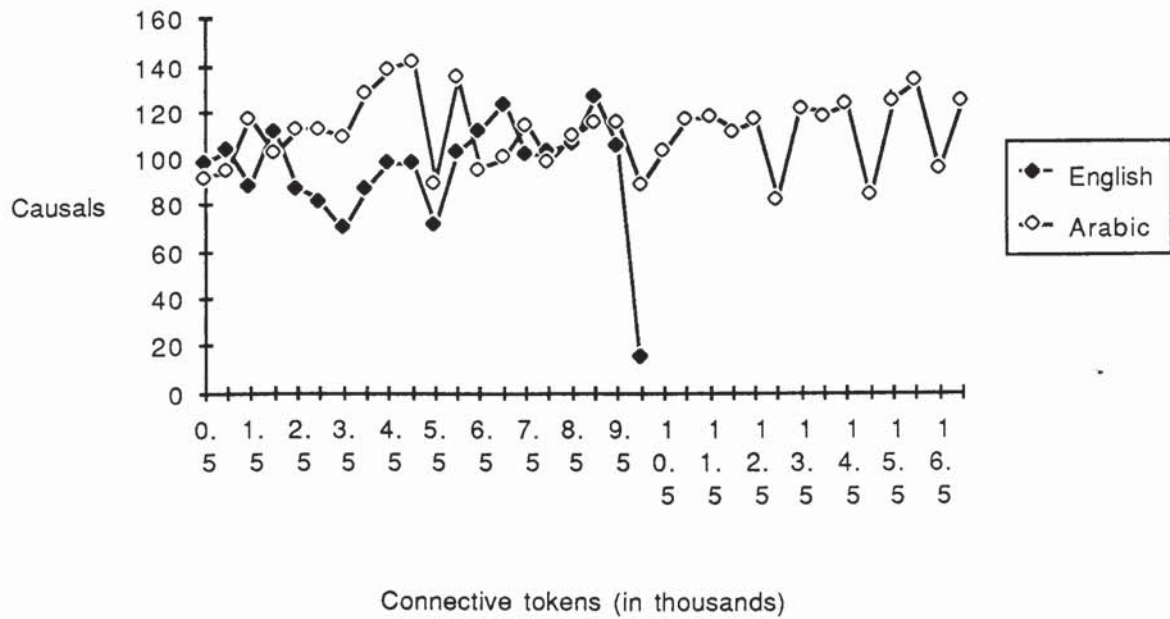


Fig. 9.55 Comparison of the Local Growth of Causal Connectives



25%, 50%, 75% and 90% of token coverage are achieved and found that they occur in the same intervals in each corpus: respectively 13 (65,000 words), 27 (135,000 words), 39 (195,000 words) and 47 (235,000 words). Extrapolation to a million word English corpus is expected to contain 7.167 causal tokens, while an Arabic corpus of a similar size is expected to contain 15,341.

ii. There are clear differences in the manner of growth of types across the two corpora. Growth starts faster in English than in Arabic. For instance, the 50th type percentile occurs within interval 4 (20,000 words) in English and 8 (40,000 words) in Arabic. After this, growth of types in English slows down to a considerable extent, while it speeds up in Arabic. For instance, the 75th type percentile occurs within interval 15 (75,000 words) in English and 14 (70,000 words in Arabic); and the 90th type percentile occurs within interval 34 (170,000 words) in English and 23 (115,000 words) in Arabic.

According to this manner of growth, an English corpus of a million words is expected to contain 66 types (a growth of 21 types). An Arabic corpus of a similar length is expected to contain 158 tokens (an increase of 82 types).

b. Local Growth

Local growth of causals is consistent in both corpora. In English, the intervals where 25%, 50%, 75% and 90% of causal token coverage are achieved are systematically spaced: 5, 11, 15 and 18. In Arabic, there is similar systematic growth, although the intervals are far more widely spaced than English. The achievement

of 25%, 50%, 75% and 90% of causal token coverage occurs within intervals 9, 17, 26 and 31 respectively.

According to this manner of growth, extrapolation to an English corpus containing 100,000 corpus is expected to include 18,441 causals. An Arabic corpus containing a similar number of connective tokens is expected to have 24,131.

4. Categories of Causal Connectives

Comparison of the size of the five categories of causal connectives in both corpora is plotted in Figures (9.56-57) and is summarised below.

a. Cause-Reason

This is a larger category in Arabic in terms of absolute frequency, comprising 1,259 tokens compared to 484 in English. In terms of relative frequency, the Arabic connectives constitute 33.3 of causal tokens while their English counterparts stand for 26%.

There are 33 types signalling this relation in Arabic, of which 5 are hapaxes. In English, the category includes 16 types, of which 6 are hapaxes. Accordingly, variegation is higher in English (37.5) relative to Arabic (15), while concentration is closely similar (0.02 compared to 0.022 in Arabic). TTR is 0.033 in English and 0.026 in Arabic.

Observation of the behaviour of the top most frequent connectives indicates that the category is dominated by 3 connectives in each language. In English, the connectives "because" (51% of tokens), "for" (19%) and "since" (12%) have a combined frequency that

Fig. 9.56 Comparison of the Distribution of the Categories of Causal Connectives in the two Corpora

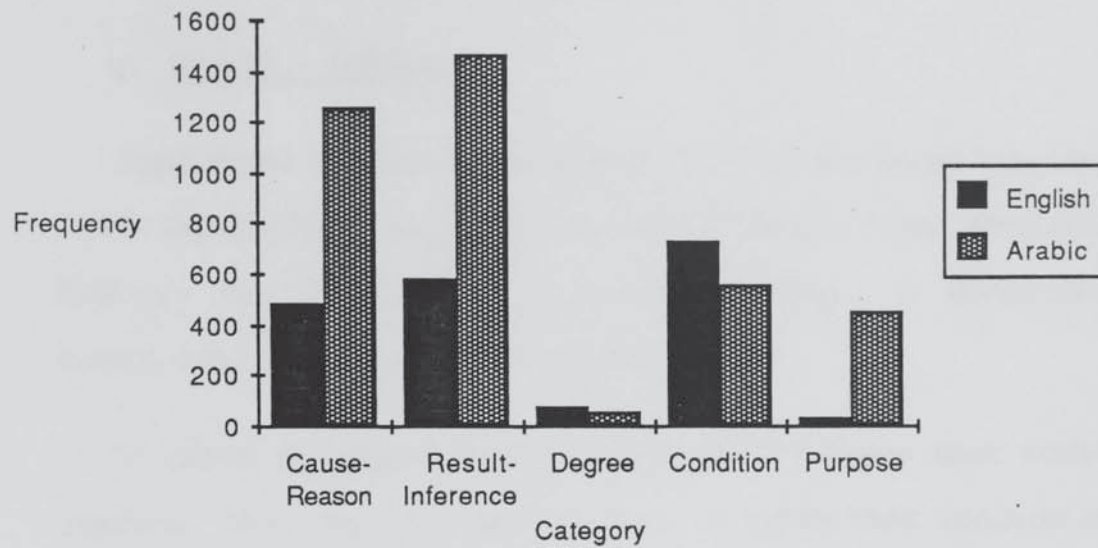
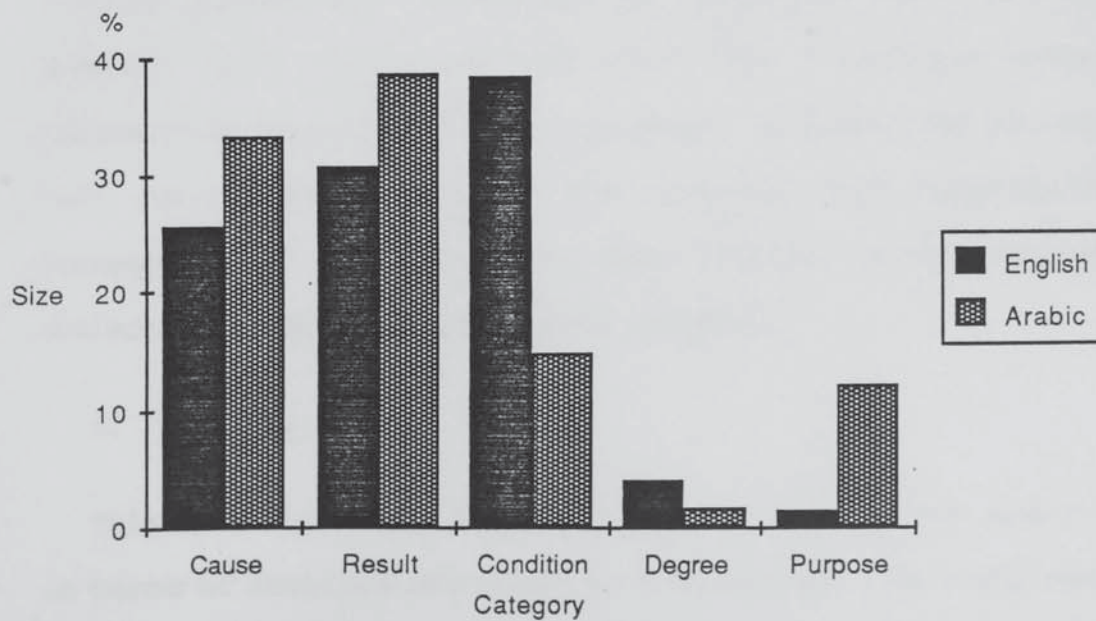


Fig. 9.57 Comparison of the Distribution of Shares of Causal Categories



represents 82% of tokens in the category. Similarly, in Arabic, connectives such as "fa" (51%), "li'anna" (23%) and "ḥaytu" (4%) achieve collectively 78% of token coverage.

b. Result - Inference

Connectives of result-inference have an absolute frequency of 1,460 representing approximately 39% of causal token coverage. In English, the category has a lower frequency. It comprises 581 tokens that stand for 27% of causals.

In terms of types, Arabic comprises 33 types that contain 5 hapaxes. English, in contrast, has 18 types that include only 2 hapaxes. TTR is higher in English (0.03 compared to 0.023), mainly due to the lower number of tokens. Concentration is higher in English, 0.027 vs. 0.017, while variegation is lower, 11 vs. 17.

In both corpora the category is dominated by the first 4 most frequent connectives. In English the connectives "and" (in a causal meaning), "so", "therefore" and "thus" have a combined frequency representing approximately 77% of tokens. In Arabic the connectives "wa" (in a causal meaning), "fa" (result) "fa" (introducing a consequence of a condition) and "ḡālika 'anna" represent collectively 69% of tokens in this category.

c. Condition

This is a noticeably larger category in English than Arabic both in terms of absolute and relative frequencies. It comprises 724 tokens in English, corresponding to 38.5% of causals. In contrast,

the Arabic corpus contains 554 tokens representing approximately 15% of causals.

The number of types is 11 in English, including 3 hapaxes. In Arabic, the number is 12, including 2 hapaxes. These close figures indicate higher TTR and concentration in Arabic (respectively 0.022 and 0.018) than in English (respectively 0.015 and 0.011). However, the index of variegation is lower: 17 compared to 27 in English.

The category, as observed in the English corpus, is dominated by the connective "if" (91% of occurrences). This high frequency is achieved by approximately 6 types in Arabic.

d. Degree - Magnitude

Although this is a relatively smaller category in both corpora, it is larger in English. It comprises 75 tokens that achieve 4% of causal coverage, while in Arabic it contains 55 tokens that correspond to 1.5% of causals. The number of observed types in the English corpus is 4 (no hapaxes) while it is 6 in Arabic (including 1 hapax).

In English the category is dominated by the connective "so ... that", which has a relative frequency of 71%. In Arabic, the top most frequent connective achieves a relatively lower token coverage: 49%.

e. Purpose

Connectives signalling the causal function of purpose constitute a significantly larger category in Arabic, comprising 450 tokens and

13 types compared to 24 tokens and 3 types in English. In terms of relative frequency, the category represents 12% of causals in Arabic and 1.3% in English.

9.2.9.2 Textual Variations of Causal Connectives

9.2.9.2.1 Cause - Reason

There are two main variations in the operationality of connectives of this category. The first one concerns the strong tendency in Arabic for overt signalling of explanation; the second considers the role of the causal connective "fa" in paragraph development.

1. Signalling of Explanation

As discussed in 6.11 connectives that introduce a cause may signal one of the two functions: either cause proper, or explanation. In Arabic there is a strong tendency for overt signalling of explanation. This is evident in the larger number of occurrences of causal/explanatory connectives in the Arabic corpus as compared to English. Whenever the content of the next stretch of text provides explanation or justification drawn from the textual world or world knowledge, a connective is expected to signify the operation. The absence of an overt signal is likely to create unnecessary processing load through extraneous activities of checking and pattern matching in order to establish the appropriate link. In other words, the absence of the appropriate connectives will create a gap in the textual flow of propositions that is likely to cause a temporary discontinuity.

2. Organisational Force of "fa"

Arabic connectives that introduce explanation are mainly "id", "dalika ('anna)", "haytu" and "fa". Out of these, the most frequently used connective is "fa". This connective plays a significant role in establishing one of the frequent paragraph patterns in Arabic.⁽²⁾ The pattern normally starts with a topical sentence that consists of a clause or a clause complex and that makes a factual statement, an assertion of opinion or a statement of an event or course of action. A causal/explanatory "fa" is then used in order to introduce a statement (or set of statements) that explores the topical sentence, offers an explanation or justification, and assists in moving out from the topic idea. This helps to achieve a topical/conceptual unity between the topical sentence and the rest of the paragraph and is an important factor in textual expansion. The following are two examples of this pattern.

- [9.44] 'inna al-ḥalaqāti al-<ilmiyyata al-<āliyyata tataḥammalū <ib'an haḍāriyyan muḍā<afan fī mujtama<i-nā al-<arabiyyi muqāranatan bi-mā yajrī fī al-<ālamī al-tāliti. fa mujtama<u-nā al-<arabiyyu yuwājihu taḥaddiyātin xārijiyyatan maṣīriyyatan 'iḍāfatan 'ilā al-taḥaddiyāti al-dāxiliyyati wa 'ahammu-hā al-taxallufu.

[The highly qualified scientific circles bear a double cultural responsibility within the Arab society in comparison to the Third World. For the Arab society faces serious external challenges in addition to the internal challenges the most prominent of which is backwardness.]

(J, 11/1/83, X76, 14663-9)

- [9.45] wa qad kānat al-<ilāqātu bayna al-duwali fī al-qurūni al-māḍiyati ta<tamidu <alā al-quwwati al-gāsimati. fa al-dawlatu allatī hiya 'aqwā jayṣan wa 'aḥsanu taslīḥan kānat tafriḍu ra'ya-hā <alā al-dawlati al-da<īfati ...

[International relations during the past few centuries were governed by power. For the nation that had a stronger and better equipped army used to impose its will on the weaker nation ...]

(Sh, 22/1/83, X108, 20980-4)

Occasionally a paragraph may require a series of linkages with causal/explanatory "fa". This is particularly true when a paragraph includes, in its expansion and development, a number of assertions that need to be elaborated or explained. Paragraph development not only tolerates but necessitates the multiple use of "fa" to achieve the requirement of unity and intelligibility and, in this way, promote textual efficiency. The following is an example:

[9.46] kullu hurriyyatin bi-lā qānūnin tanqalibu 'ilā fawḍā ... Fa 'ayyan kāna naw<u al-ḥurriyyati lā budda wa 'an yuṣāhibu-hā ba<ḍu al-dawābiṭi wa allatī lā yakūnu hunāka buddun min 'an takūna ... wāḍiḥatan ... Fa mata mā 'axaḍat al-ḍawābiṭu bi-<ayni al-'<tibari <adama al-'inqāṣi min al-ḥaqqi al-makfūli wa kānat wāḍiḥatan lā gubāra wa lā labsa fī-hā, kafalat ḥurriyyāti al-'afrādi ...

[Freedom without a law that regulates it will turn into chaos ... For whatever kind of freedom it is, it should be accompanied by a number of regulating restrictions that have to be made clear ... For when the restrictions do not affect guaranteed rights and are rendered clear, with no ambiguities or misinterpretations, they will secure freedom for the individuals.]

(Nb, 15/7/83, X21, 4128-42)

9.2.9.2.2 Result - Inference

Observation of the behaviour of connectives signalling a relation of result and inference in both corpora suggests two main variations. Both are related to the nature of the relation: one examines the strictness as opposed to vagueness of signalling; the other looks into the conflation of additivity and temporality with causality.

1. Strictness and Vagueness in Signalling

Connectives that introduce a statement of result or consequence operate on a broad continuum. At one end, the relation is strictly causal, i.e. the content of the next stretch of text (the subsequent) represents a clear consequence of the content of the current one. At the other, the relation is only vaguely causal, i.e. the content of the next stretch of text is not directly perceived as an expression of a results/consequence; more commonly, they express a comment, or an assertion that is determined by the content of the current stretch of text. In signalling the relation, connectives of both languages may occur on any point of the continuum. However, there is a clustering that represents tendencies in signalling in both languages. English connectives

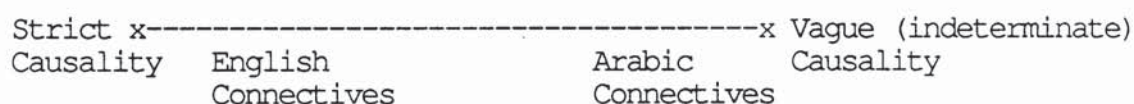


Fig. 9.58 Clustering of connectives introducing result/
consequence on the relation continuum

tend to cluster near the strict end, while Arabic connectives tend to cluster at the other end (Figure 9.58). In other words, there is a strong tendency for English connectives to signal a clearly conceived relation of causality. In contrast, Arabic connectives are more likely to signal indeterminate and remotely related causality.

Part of the indeterminateness of causality in Arabic is related

to the frequent use of the connectives "wa" and "fa" to signal a causal relationship. These two connectives conflate an additive and temporal/sequential relations with causality, producing a complex amalgam of shades of meaning that subdues the expression of causality. This is evident in the way "fa" operates in the following typical examples.

- [9.47] ... ya<mal-ūna bi-'ijūrin 'aqalla wa juhḍin
'akbara. fa yatimmu ta<wiḍu al-xasā'iri.
[... they put more effort for lesser wages. And so
all losses can be compensated for.]
(Hr, 15/4/83, X48, 8897-8)

- [9.48] ... zādat fa 'aṣḥaḥat 'alfa ma<hadin.
[... they [the academic institutes] were expanded
and as a result (until, and so) their number
reached a thousand].
(Hr, 4/1/83, X25, 5344-5)

Indeterminateness of causality may sometimes stem up from the explanatory force of the connective "fa". This happens when, in introducing an explanation, ambiguity arises as whether "fa" introduces a cause (that is, it is equivalent to "for") or a result/consequence (equivalent to "so" or "then"). For example:

- [9.49] ... yastaṭī<u al-tilifizyūnu 'an yasudda al-ṭagrata
al-ṭāri'ata al-nāṣi'ata <an bawāri al-kitābi fī al-
waḳti al-rāhini. fa ya<riḍu al-kitāba wa yuwajjihu
al-naẓara 'ilay-hi wa yunaṣṣiṭu tadāwila-hu.
[... Television can fill the temporary gap that is
created by of books. Accordingly (For) it can
exhibit a book, draws attention to it and expands
its circulation.]
(Hr, 18/4/83, X51, 9456-60)

It should, however, be noted that Arabic has, within this category, a class of connectives that tend to signal a stricter expression of causality. This is the group of connectives that

introduce inference or conclusion, particularly "'idan" and certain uses of "bi-al-tāli", "liḏā", "li-hādā" and "li-dālika".

2. Introducing questions

The connective "fa" may be used to establish a rhetorical pattern that is frequent in Arabic argumentative text. The pattern involves one or more assertions that establish a viewpoint. To consolidate them, the text producer introduces one or, more often, a set of questions that are related to and, more frequently, derived from, the viewpoint. These questions are then used to build up the argument either through expansion of the viewpoint or the establishment of another. Since Arabic is sensitive to discontinuity, rhetorical unity within the pattern is sustained by using "fa" to link the question (of the first one in the case of a set of questions) with the preceding assertion(s). Exemplification requires lengthy citations to establish the pattern; we shall therefore give only part of the full quotations.

- [9.50] ... wa šārat al-mintāqatu miṭālan fī al-tamazzuqi wa al-fawḏā wa al-'inhizāmiyyati ... fa hal tammata 'amalun fī 'iḳādati al-naẓari bi-ru'yatin šāfiyatin tatajāwazu al-'ananiyyata wa al-'ahqāda? wa hal tammata 'amalun fī 'inqādi mā yumkinu 'inqādu-hu ...? [... the region has become an example of disintegration, chaos and escapism ... so is there a hope of a clearly-envisaged reassessment that may disregard selfishness and hatred? And is there a hope of saving what can be saved?...] (Hr, 31/3/83, X42, 7765-72)

- [9.51] ... 'inna 'isrā'ila nafsa-hā qad (istafādat) min al-tawrati al-jazā'iriyyati ... fa kayfa kāna ḏālika? [... Israel made use of (the lessons of) the Algerian war ... (So) how did that happen? (how can that be explained?)]

(Ar, 13/6/83, X205, 35365-9)

In English, such a pattern is available, but to a lesser degree of frequency. But even when it is available, a link via a connective (usually "so") is possible but not necessary as in Arabic.

9.2.9.2.3 Condition

A number of textual variations in signalling condition have already been discussed in 6.11.6 above. We shall examine below two further differences: one concerns certain constraints on the surface behaviour of Arabic connectives that have not been observed in English. The second is related to the general functionality of the connective "if" in comparison to the specific functionality of the Arabic equivalent connectives.

1. Constraints on Surface Behaviour

Arabic connectives of condition are constrained in their positioning and their association with other connectives. The normal position for conditional statements is one where statement B (the consequence) precedes statement A (the condition). Since the connective is attached to statement A, its position is medial in the sequence. In the less frequent cases where statement A precedes statement B and, therefore, the connective is in initial position in the sequence, the conditional statement suffers from textual discontinuity. To neutralise this factor and so sustain text stability, the conditional statement is supported via connective association of two types:

- a. The conditional statement is related to the previous text

sequence by a supportive connective, particularly "wa" and "fa". The connective immediately precedes, and is therefore associated with, the conditional connective. This association can be bypassed in very restricted cases: when the conditional statement is i) at the start of the text, i.e. the first statement in the text; ii) at the start of a major text segment, such as one after a subtitle; or iii) at the start of a quotation. Cases (i) and (ii) do not occur in the corpus; case (iii) occurs with a very low frequency.

b. The conditional statement is related to statement B (the consequence) with i) the connective "fa", which has in this case a meaning similar to "then" in English; or ii) the particle "la", particularly when the conditional connective is "law".

2. General vs. Specific Functionality

In English, the most frequent signal of condition is the connective "if". This connective is capable of expressing a multiplicity of conditional meanings and is therefore a signal for general functionality. To match this complexity of signalling, Arabic utilises a number of connectives, each signifying its specific conditional meaning. The most frequent of such connectives are "'idā", "law" and "'in". The connective "'idā" combines condition with temporality.⁽³⁾ It has been defined by Arab grammarians as "an adverb signifying future time". The connective "law" is most frequently used to express hypothetical condition, particularly when referring to past time. The connective that expresses pure conditionality (i.e. temporality is not necessary) is "'in". The functional force of this connective affects the mood of

both verbs in the subordinate and main clauses, turning them into the jussive. But this connective is less frequent in the corpus, having a frequency of 50 (9% of conditionals), than "law" (69 occurrences, 12.5% of conditionals) or "'idā" (338 occurrences, 69% of conditionals). The higher frequency of the latter is indicative of a tendency in Arabic to conflate conditional and temporal meanings, which has the effect of "toning down" the functional force of the relation.

3. Positioning and Functionality

Another variation that we shall include here but will not elaborate further, mainly because it requires a more detailed investigation, is the influence of positioning and arrangement of statements A and B on functionality. Initial observation of positioning of "'idā" in Arabic suggests that if statement A precedes statement B (which is not the frequent arrangement in a sequence, see point 1 above), the conditional meaning is more likely to be "rhetorical" (for discussion of rhetorical condition see 6.11.6). A similar constraint has not been observed in the positioning of "if" in English.

9.2.9.2.4 Degree - Magnitude

The main variation in the behaviour of connectives of this category in the two corpora is the relative flexibility of the English connectives compared to their Arabic counterpart. English connectives are capable of directing focus to a target concept within the antecedent and thus relating the rest of the sequence to it. Arabic connectives display a relative inability to exercise a

similar task.

To elaborate this, we should note that Arabic does not have an equivalent connective to "so ... that" or "such ... that" that can arrange the propositions and exploit their conceptual content with the same degree of efficiency. These two connectives are capable, among other things, of directing focus to a particular concept within statement A (the antecedent) by being placed on the surface structure in close vicinity to the expression that denotes that concept. In this position the semantic space of the concept is intensified to an extent where the content of statement B is directly related as a consequence. Further, the high frequency of "so ... that" and "such ... that", standing for 88% of causal connectives of degree, makes their textual pattern representative of the functionality of this category.

Arabic connectives, in contrast, looks back at the entire conceptual content of the antecedent (statement A) intensifying it to an extent where statement B becomes a compatible consequence. This pattern is represented by the connective "ḥattā". For example,

- [9.52] wa al-wāqī<u 'anna 'isrā'īla munḍu al-'iḥtilālī
'ittaxaḍat min janūbi lubnāna sūqan tijāriyyatan
maftūḥatan, 'ittasa<a li-yakūna sūqan li-al-
tabāduli al-siyāhiyyi wa sūqan li-al-<ilāji al-
ṭibbiyyi ḥattā 'aṣbaḥa min ḥaqqi kulli lubnāniyyin
'an yu<ālaja fī al-mustaṣfiyāti al-'isrā'iliyyati.
[The truth is that Israel, since its occupation of
the Lebanon, has turned the South (of Lebanon) into
an open commercial market which has expanded to
include tourism and medical treatment to the extent
that the Lebanese have the right to seek medical
treatment in the Israeli hospitals.]
(Sh, 26/6/83, X146, 28155-60)

In this excerpt, "ḥattā" initiates operations of regression where the content of the current statements is reconsidered and intensified in order to act as a cause for the consequence expressed in statement B.

9.2.9.2.5 Purpose

There are two variations that characterise the patterning of Arabic in comparison to English connectives of purpose. Both variations are local in the sense that their impact does not go beyond the sequence where the connectives operate. The first variation concerns the imposition of some structural constraints on statement B while the second is related to the problem of discontinuity that characterises the functioning of certain connectives.

1. Structural Constraints

The Arabic connectives "li", "ḥattā", "likay", "kay", "kaymā" and "likayma", the frequency of which represents 96% of connectives of this category, require that a) the subsequent statement should be a verbal clause; b) the verb is in the imperfect tense; c) the verb is marked in the subjunctive mood. The following is a typical example

[9.53] 'inna tajribata al-yābān talfitu-nā 'ilā 'umūrin lā
budda min al-taxalluṣi min-hā ḥattā nabda'a ḥayātan
jadīdatan ...

[The Japanese experience draws our attention to
certain weaknesses that we have to abandon in order
that we may start a new well developed life ...]

(Hr, 23/6/83, X63, 11395-7)

The clause that the connective introduces is verbal, i.e. having a verb in a thematic position, and the verb "nabda'a" is imperfect in

a subjunctive mood.

2. Problem of Discontinuity

Discontinuity arises when a) statement B precedes statement A so that the connective is in initial position in the sequence; b) statement A is a nominal clause. In such cases, the two statements are disconnected despite the existence of the connective. Connection is retrieved by inserting the connective "fa". For instance, in the following excerpt:

[9.54] wa ḥattā lā yakūna hunāka 'ayyu labsin fa 'inna
'amtāran qalīlatan qad takūnu 'aktara 'ahammīyyatan
min mi'ātin min al-kilūmatrāti 'idā kāna la-hā
mawqī'un 'istrātījiyyun hāmmun.

[And in order that we avoid any misunderstanding,
(we should state that) a few metres (of land) are
more important than hundreds of kilometres if those
metres have significant strategic position.]

(Hr, 26/1/83, X35, 6691-4)

the connective "fa" is used to preserve connectivity between the two statements.

9.2.10 Adversative Connectives

9.2.10.1 Quantitative Variations of Adversative Connectives

1. Frequency

a. This is the largest category in English in terms of absolute and relative frequencies. It comprises 2,404 tokens that represent 25% of total connective tokens. In Arabic the category consists of 1,692 tokens that represent 10% of total connective tokens. Figure (9.59) displays a comparison of the distribution of adversative

Fig. 9.59 Comparison of the Distribution of Adversative Connectives

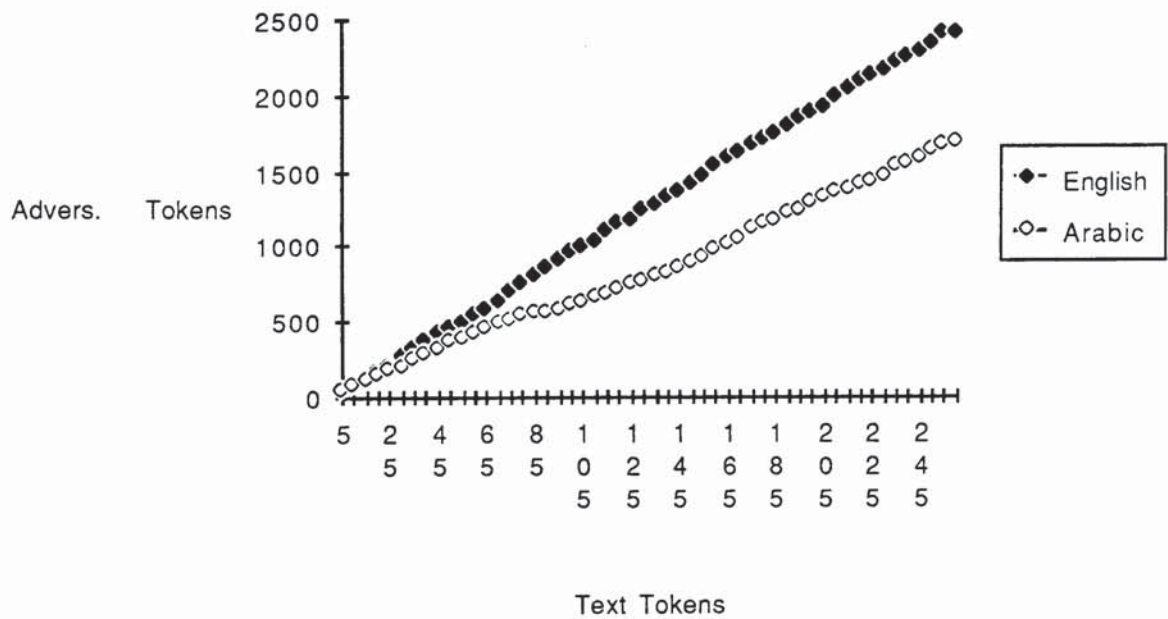
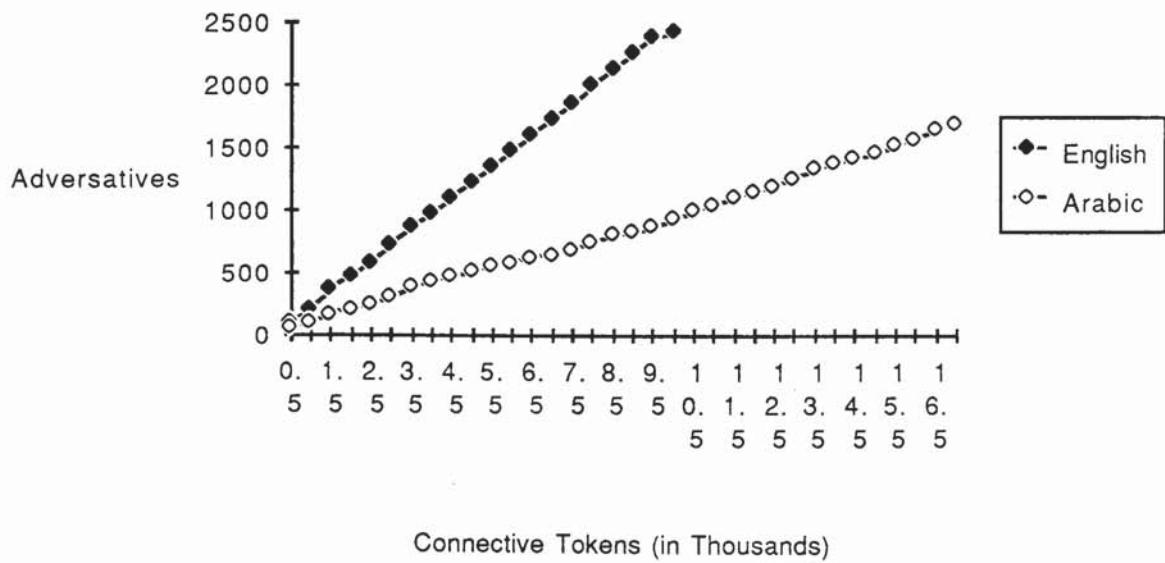


Fig. 9.60 Comparison of the Distribution of Adversatives within Connective Tokens



connectives in the two corpora. Figure (9.60) plots the distribution of adversative tokens against connective tokens.

b. In the English corpus, adversative connectives comprise a higher number of types than their counterparts in Arabic: 72 vs. 53. The English set includes 19 hapaxes while the Arabic set has 8. These figures suggest slightly higher TTR and concentration in Arabic (respectively 0.032 vs. 0.03 and 0.027 vs. 0.022 in English). However, the category in English manifests higher variegation in the use of types (26 vs. 15), higher gravity (126.5 vs. 211.5), rhythmicity (44 vs. 36), stereotypicality (45 vs. 37) and density (11.3207 vs. 10.7755).

c. In English the category is dominated by 5 connectives whose combined frequency constitutes 70% of adversative tokens. These are "but" (48.5%), "however" (8%), "though" (5%), "although" (4.2%) and "yet" (4.2%). In Arabic, a similar frequency is achieved by 9 connectives (details are given in Appendix 140).

2. Repetitiveness

Adversative connectives in the English corpus manifest higher rates of repetitiveness than their counterparts do in Arabic. This is evident in the various indices that are calculated. For instance, adversative occurrence rate is 4 in English and 10 in Arabic; general repeat rate is 0.06 vs. 0.01 in Arabic; and adversative system repeat rate is (159×10^{-4}) in English and (9×10^{-4}) in Arabic. Average length of distances separating two occurrences of adversative tokens is 106 words in English and 150 words in Arabic. Type occurrence rate is slightly higher in Arabic

31 compared to 33 in English.

3. Growth

Comparisons of global and local growth are plotted in Figures (9.61-62) respectively.

a. Global Growth

i. The pattern of growth of adversative connectives in each corpus differs from each other. This is reflected in the growth rate, which is calculated, using Tuldava's formula, as -5.0425 in English and -3.0784 in Arabic. Examined at various intervals, growth manifests the following variations: a) Arabic achieves the first 25% of adversative tokens faster than English: this point occurs within interval 12 (6,000 words) in Arabic and 14 (7,000 words) in English. b) The next 25% is achieved at a much slower growth rate in Arabic than English. The 50th percentile point occurs within interval 30 (150,000 words) in Arabic and 26 (130,000 words) in English. c) Growth then moves slightly faster in Arabic; the 75th percentile point occurs within interval 38 (190,000 words) in English and 40 (200,000 words) in Arabic, and the 90th percentile point occurs within interval 46 (230,000 words) in English and 47 (235,000 words) in Arabic.

According to the pattern and size of growth, a million word English corpus is expected to contain 10,082 adversative tokens. An Arabic corpus of a similar size is expected to contain 5,398.

ii. Although growth of types starts at an equal rate (both

Fig. 9.61 Comparison of "Global" Growth of Adversative Connectives

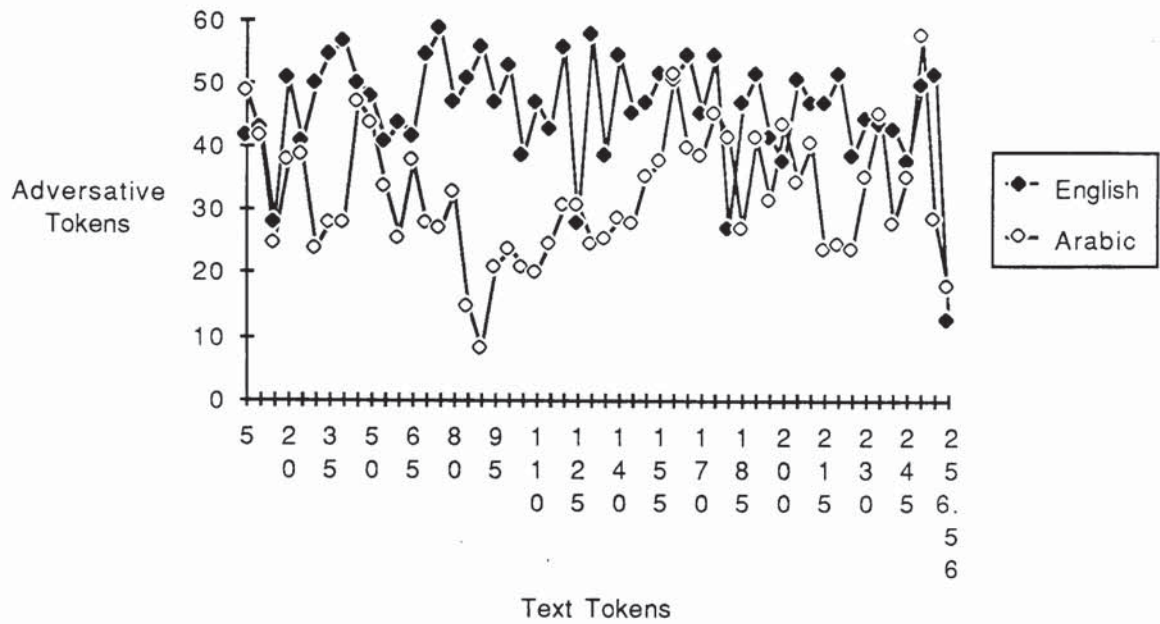
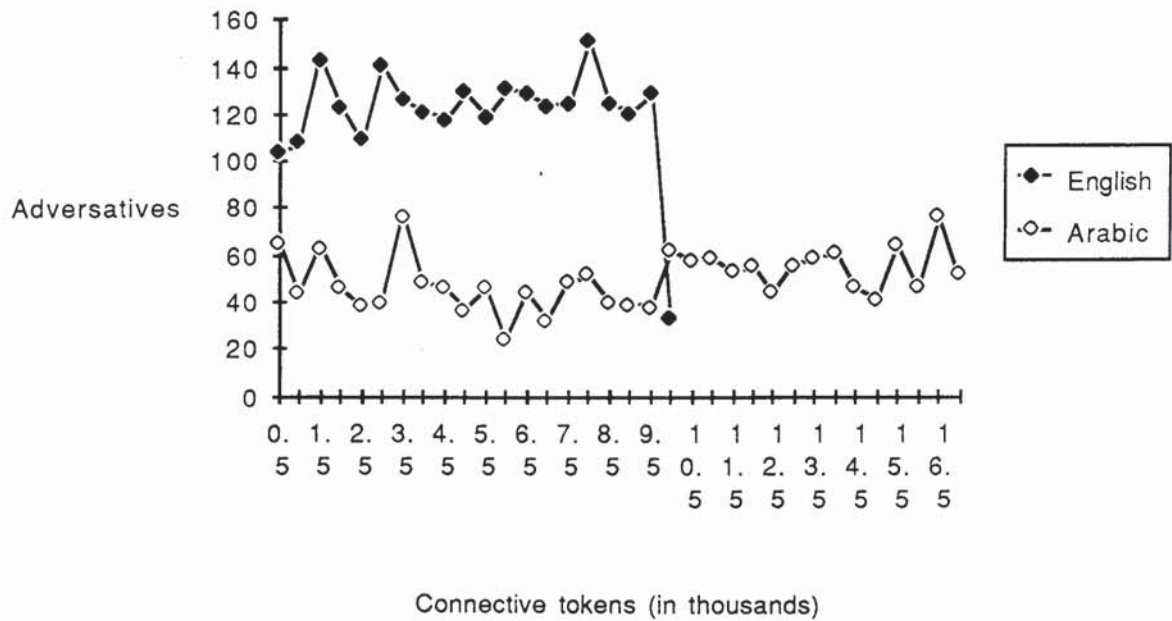


Fig. 9.62 Comparison of the Local Growth of Adversative Connectives



corpora achieve 25% of types within the first interval, i.e. the first 5,000 words), it moves on at a faster rate within the next few intervals. The 50th type percentile occurs within interval 5 (25,000 words) in Arabic, while it occurs within interval 7 (35,000 words) in English. The 75th percentile type occurs within interval 14 (70,000 words) in Arabic and interval 18 (90,000 words) in English. After that, growth slows down considerably in Arabic relative to English, and thus signs of saturation start to appear earlier in English, with the 90th percentile type occurring within interval 30 (150,000 words) in English while it occurs within interval 34 (170,000 words) in Arabic.

According to this pattern and size of growth, an English corpus of a million words is expected to contain 127 types. An Arabic corpus of a similar length is expected to include 82.

b. Local Growth

One immediately discernible variation is the ratio of adversative connectives to the size of interval (which is 500 connective tokens). This ratio is consistently higher in English than in Arabic. For instance, the maximum ratio in English is 0.3 at interval 16 (8,000 connective tokens), while in Arabic it is 0.15. The minimum ratio (excluding the last interval which is of a smaller size) is 0.21 in English and 0.05 in Arabic. This is due to two factors: the larger size of the category in English and the higher number of connective tokens in Arabic. According to the pattern of growth, extrapolation to an English corpus comprising 100,000 connective tokens is expected to contain 27,634 adversative tokens.

Extrapolation to an Arabic corpus comprising a similar number of connective tokens is expected to include 8,403.

4. Categories of Adversative Connectives

Comparison of the absolute and relative frequencies of the two adversative categories across the two corpora are plotted in Figures (9.63-64) and is summarised below.

a. Antitheticity

Antithetic connectives comprise a larger set in the English corpus. It is made up of 2,149 tokens that represent 89% of adversatives. In Arabic, the set comprises 1,359 tokens that correspond to 80% of adversatives. Further, the English set contains 49 types including 12 hapaxes, while the Arabic set has 37 including 7 hapaxes.

According to these figures, Arabic adversatives have slightly higher TTR (0.027 compared to 0.022) and a similar concentration (0.022 in both corpora). However, English adversatives display higher variegation in the use of types (24.5 compared to 19 in Arabic), higher gravity (179 compared to 194) and higher repeat rate among adversatives (0.8 compared to 0.65).

In English, the distribution of antithetic connectives is dominated by the frequency of the first two connectives, which collectively constitute 63% of connectives in the category. These are the connectives "but" (54%) and "however" (9%). In Arabic, a similar share of tokens is achieved by 5 connectives (refer to the details in Appendix 145A).

Fig. 9.63 Comparison of the Distribution of the Categories of Adversative Connectives in the two Corpora

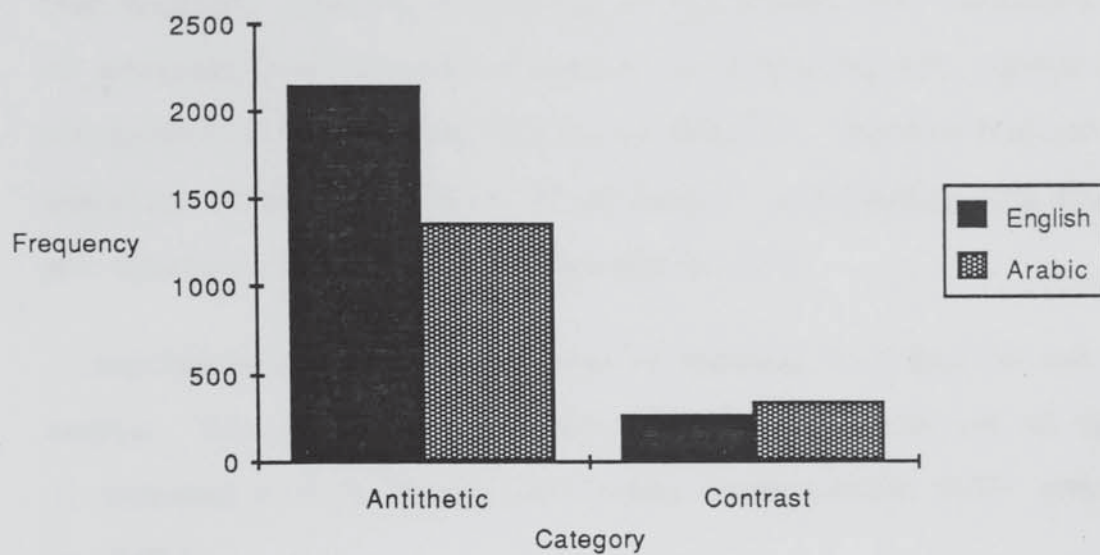
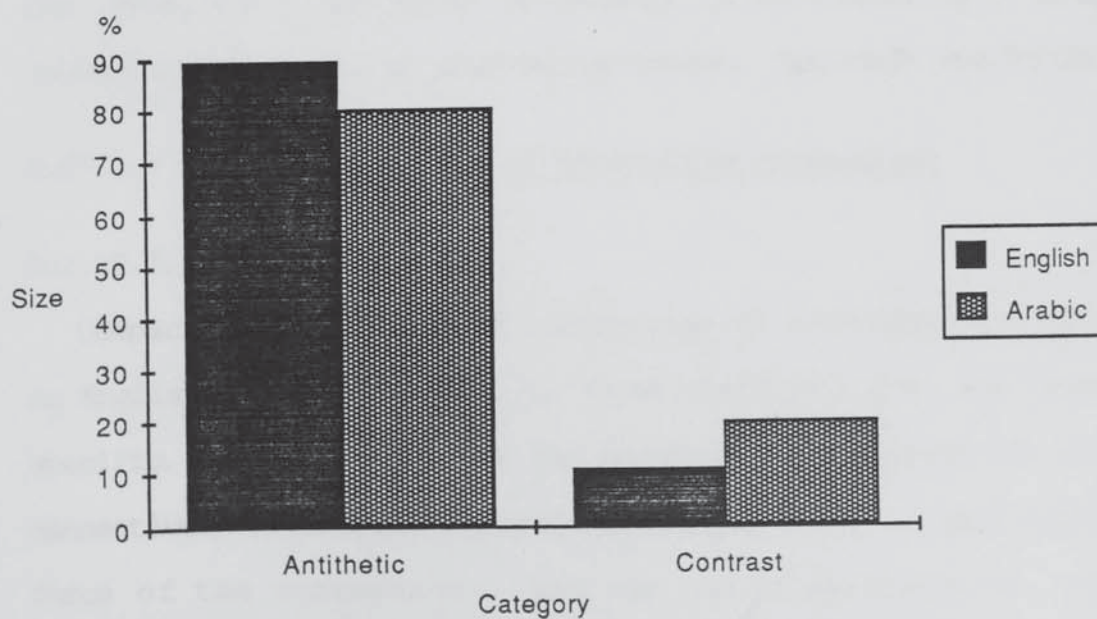


Fig. 9.64 Comparison of the Distribution of Shares of Adversative Categories



b. Contrast

Connectives of contrast constitute a larger category in Arabic than English. The set is made up of 333 tokens that represent 20% of adversatives in Arabic, while it comprises 255 tokens that correspond to 11% of adversatives in English. But the English set comprises more types: 26 vs. 21 in Arabic. Accordingly, the English set displays higher TTR, 0.1 compared to 0.06.

Another variation is the number of hapaxes: 8 in English and 2 in Arabic. This gives English higher variegation in the use of types: 31 compared to 9 in Arabic, and higher concentration (0.071 compared to 0.057).

In both corpora, the frequency distribution is dominated by the top most frequent connectives. For instance, in English the connectives "but" and "instead" constitute 51% of token coverage in the category. The same frequency is achieved by 3 Arabic connectives "wa" (in an adversative sense), "baynamā" and "illā".

9.2.10.2 Textual Variations of Adversative Connectives

9.2.10.2.1 Antitheticity

Comparison of the textual patternings of antithetic connectives in English and Arabic points to three variations that are language specific. The first concerns the requirement of support for Arabic connectives, the second concerns variation in the organisational force of the connectives, and the third pertains to some peculiarities of patterning. These are discussed in more detail below.

1. Requirement of support

One striking variation in the behaviour of connectives is the need of Arabic connectives for support in order to function cohesively in the text. This support is provided via the utilisation of the connective "wa", which precedes the connective and acts to eliminate possibilities of discontinuity. The absence of this support leads to disintegration in the unity of the segment and may have adverse effects on processing operations.

The connectives that are usually supported via additivity wherever their position is in the text sequence are "lākinna", "lākin" and "'innamā". The combined frequency of these three connectives constitutes approximately 43% of antithetic connectives in the corpus. Connectives that are supported when they introduce the antecedent, i.e. occur initially in the sequence, are "ragma 'anna", "ma<a 'anna", "sawā'an" and "mahmā".

There are, however, five connectives that do not require additive support. Two have already "wa" as an integral part of the structure of the connective: "wa 'in" and "wa law". The rest are "gayra 'anna", "'illā 'anna", and "bal". All five connectives are immobile, i.e. their position in the sequence is fixed and therefore they cannot be commuted.

2. Organisational Force

Inspection of the organisational force of antithetic connectives in both corpora indicates that Arabic has only a small number of

equivalents to the English connectives "however" and "yet". These two connectives are capable of relating sequences of various sizes to each other. They can connect antagonistically related sequences within one paragraph or across two or more paragraphs. Hence they are useful tools for developing intra- or inter-paragraph structure and organising propositions within the textual world.

Arabic antithetic connectives, in contrast, have a shorter range of application. That is, their organisational force is limited to relating relatively short text sequences and, therefore, their main field of operation is within the paragraph. In the few cases where certain antithetic connectives operate across the paragraph, the operation is assisted by another connective, usually "wa" or "fa", which precedes the connective and establishes the link.

3. Peculiarities of Patterning

Some connectives in Arabic have patterns of operations that are peculiar to them. These patterns take the form of constraints that are either textual or structural.

a. "'innamā" and "bal": Both require that the content of the antecedent is negated. This enhances the alternative that "'innama" introduces. For example:

[9.55] wa las-nā <ālamān tālīṭān wa 'innamā nahnu
 juz'un min hādā al-<ālamī al-wāḥidī.
 [We are not (by ourselves) a third world, but we
 are part of one single unified world.]
 (Nb, 4/4/83, X2, 185-6)

Note that "bal" achieves a sharper opposition between the contents of the two statements than "'innama".

b. "lākinna" and "lākin": The connective "lākinna" introduces nominal clauses only. But this constraint is overruled when the connective is changed into "lākin" (by reducing the geminated "nūn" into a single one). The connective "lākin", unlike "lākinna", can introduce nominal or verbal clauses.

9.2.10.2.2 Contrast

Inspection of the patterning of connectives of contrast points to two noticeable areas of difference. One concerns positioning while the other is related to the Arabic use of additivity in a contrastive environment. The direction of examination is from Arabic to English.

1. Constraints in Positioning

Most connectives of this category are mobile. The most frequent positions are initial or medial in the sequence. Arabic connectives, however, are constrained textually when they occur initially. In that position, a connective creates discontinuity, i.e. it disintegrates the sequence from the previous text sequences and creates a gap that is deep enough to disturb cohesion. To bridge the gap, connectivity falls back on additivity and the sequence is related to the rest of the text with the connectives "wa" and "fa" (unless another connective already exists that performs this task). The only two connectives that are immobile are "'ammā", which is positioned initially in the sequence, and "'illā 'anna", which is positioned medially. Each of these two, however, has its own peculiarity of patterning, as will be indicated in (3)

below.

2. Use of Additivity

The constraint explained in (1) above is an evidence confirming that Arabic is more prone to using additivity in a contrastive environment than English does. Another evidence is the frequent use of "wa" to combine statements that are in contrastive opposition. This connective is basically additive; however, in this position it conflates an additive with an adversative/contrastive function. For instance, in the following excerpts:

[9.56] wa lākinna hādā al-sabqa lam yamna< min 'an takūna
'almānya min al-<ālamī al-'awwalī, wa al-hindu wa
kullu al-mujtami<īna fī <āṣimati-hā min duwalī
<adami al'inḥiyāzi yantam-ūna 'ilā al-<ālamī al-
tālīti.

[But the jet time lag did not prevent Germany from
being a first (i.e. developed) world country
whereas India and all non-aligned countries
represented in (the Conference held in) it belong
to the Third World.]

(Nb, 4/4/83, X2, 115-9)

the connective "wa" appends the two statements additively while simultaneously creating a sharp prominence by setting their conceptual content in opposition.

3. Peculiarities of Patterning

a. "'ammā": Basically this is an orientative connective, but may in certain cases signal contrast by introducing the subsequent and setting it in parallel to the antecedent. The connective requires the use of "fa" to help place the content of the subsequent in focus. For example:

[9.57] ... kāna 'itnāni min majmū'i al-mustaxdamīna al-
ṭamānīna fī 'aḥadi maṣārifi <āṣimatin min <awāṣimi
al-'aḡṭāri al-xalījiyyati min-hā; 'ammā al-bāqī
min-hum fa min al-<amālati al-wāfidati.

[... only two of the eighty employees in a bank in
one of the capitals in the Gulf are compatriots of
that state; the rest are foreign labour.]

(J, 15/1/83, X77, 15049-53)

b. "illā 'anna": The pattern of this connective normally
requires that the antecedent is in the negative. This helps
intensify the function of the connective as it introduces an
opposing or parallel statement.

9.3 Range of Operationality

9.3.1 Preliminaries

Connectives, as discussed in Chapter 4, are capable of relating
text sequences, and by doing so, of creating and organising text.
The semantic space of the connectives can encompass propositions to
create conceptual chunks or combine chunks to construct the text
world: the total knowledge activated while processing the text. The
extent to which the semantic space of a connective accommodates
conceptual configurations is here referred to as its range of
operationality.

Connectives differ in the magnitude of their operational range.
Close examination of the textual functioning and rhetorical
patterning of connectives in the two corpora has led us to
distinguish several types of range. The criteria that we have used
is:

a) The volume of text chunk involved in the operation. Some
connectives are capable of relating and combining relatively large

sequences, such as paragraphs or "paragraph clusters" (text segments involving more than one paragraph). Others operate on relatively shorter sequences.

b) The distance that separates the subsequent from the antecedent. Distance is here identified with the boundaries of text constituents. Connectives capable of bridging paragraphs, paragraph clusters or larger sequences (for instance text sections whether marked or not with subtitles, section numbers, etc.), operate across relatively long distances. Those that combine clauses within a clause complex have a relatively shorter distance to bridge.

Both volume and distance have to be understood as relative. In certain cases, a paragraph may have a smaller volume than a segment of another paragraph. Likewise, a pair of segments within a paragraph may display a longer distance than another, structurally similar, pair within a different paragraph. However, this factor should not compromise the adoption of the two criteria for analytical purposes.

9.3.2 Types of Range

The combination of the two criteria helps us to identify four types of range. For convenience, these are here labelled long, medium, short and immediate.

1. Long: A connective with a long range operates across the paragraphs and relates two large chunks of text. Such connectives normally occur initially in the subsequent paragraph, but may also occur anywhere within the first sentence in that paragraph. A long

range implies powerful operationality that relates the major components of the text and assists in sustaining its unity. If we accept that paragraphing (or larger segmentation) correlates with topical development in a text, we can then agree that relating these large segments not only organises textual development in an orderly and logical way but contributes considerably to its unity and coherence. It follows that connectives with a long range display organisational force that, when properly utilised, achieves efficiency of textual design.

2. Medium: Connectives having this range operate on two segments (for instance two sentences) within the same paragraph that, nevertheless, do not lie in close proximity. In other words, there are other segments, i.e. sentences, that separate the connected objects (the antecedent and subsequent). This range, as will be shown later, is not very common. However, its organisational force is strong as it assists in producing a tightly organised paragraph.

3. Short: Connectives having a short range operate across sentence boundaries, normally relating two sentences. Since sentence boundaries often, though not always, function as conceptual boundaries as well (cf. Goldman-Eisler 1972, Just and Carpenter 1980, Beaugrande 1984), the connective with a short operational range assists in creating larger conceptually integrated configurations of text sequences from relatively smaller ones. The integration contributes to the creation of hierarchies of conceptual content, most notably the evolution of the paragraph.

4. Immediate: Connectives that have an immediate range operate on combining clauses into clause-complexes of various types and sizes. The combination creates a scale of syntactic/conceptual dependencies between the connected objects (the antecedent and subsequent): some connectives reflect heavier dependencies than others. At one end of the scale connectivity is based on parataxis and the connected objects manifest relative independence. At the other end, connectivity utilises hypotaxis where there is, as a minimum requirement, a clear syntactic/conceptual dependency of one clause, the dependent, on another, the independent.

We should note here that occasionally connectives manifest some indeterminateness of operational range. This is particularly true when either the range is not immediately obvious and therefore is not easy to pin down, or, more frequently, when the connective includes some exophoric meaning and therefore its range covers general world knowledge in conjunction with text world knowledge. We have, in such cases, allocated the connective the most approximate range of operationality to the main four types discussed above.

We now turn to a consideration of the distribution of each type of operational range of connectivity in each corpus and examine their role in maintaining text organisation.

9.3.3 Range of Connectives: A General Profile

Distribution of range of operationality is displayed in Table (9.7). The following observations summarise the variations across the two corpora.

1. In general, there is a similar ordering of the frequency of range in both corpora. The most frequent range is that of immediate operational range. This is followed by the frequencies of short, long and then medium ranges. One reason that explains this type of

Range	<u>English</u>		<u>Arabic</u>	
	Abs.	Rel.	Abs.	Rel.
Immediate	6335	66.02	9717	57.17
Short	1885	19.64	4458	26.23
Medium	195	2.03	489	2.88
Long	1181	12.31	2331	13.72
Total	9596	100.00	16995	100.00

Table 9.7 Distribution of operational range of connectives

ordering is the nature of the distribution of textual sequences in each corpus: there are higher numbers of clauses than sentences, and higher numbers of sentences than paragraphs. The low frequency of medium range operatinality in each corpus can be attributed to two related factors: a) the intricacy of the regression/progression operations involved in establishing medium range connectivity, and b) the more urgent concern of connectivity towards relating sequences in closer proximity than those separated by other sequences.

2. One noticeable variation concerns the distribution of connectives used for immediate range connectivity. The number of such connectives is higher in the Arabic corpus compared to the English: 9,717 vs. 6,335. However, if we consider their relative

frequencies, then English utilises a larger share of its repertory of connectives for sustaining this type of connectivity (66% compared to 57% in Arabic).

3. If we relate the absolute frequencies mentioned in (2) above to the number of sentences in each corpus, then we can estimate that, on average, English immediate range connectives operate within, and so help to create, 54% of sentences in the corpus, assuming that each of these sentences has one immediate range connective that relates two constituent statements. In Arabic, however, such a calculation gives widely different results. This type of connectives is estimated to operate on every sentence in the corpus and, on top of that, there are 20% of sentences where 2 short range connectives are operative. This indicates the reliance of text organisation in Arabic in creating longer "chunks" (in this case, sentences) by relating an average of two or more statements (or clauses).

4. Another variation concerns short range operability. Both the absolute and relative frequencies of connectives having this type of range in Arabic are higher than in English. There are 4,458 connectives in the Arabic corpus operating across adjacent sentences, compared to 1,885 in English. These numbers represent more than 26% of connective mass in Arabic and less than 20% in English. Accordingly, there is a stronger tendency in Arabic to relate sentences via the use of connectives

5. The influence of the frequencies mentioned in (4) above is better elucidated when the absolute numbers are related to the

number of sentences in each corpus. In English, connectives are estimated to operate on 16% of sentences in the corpus, assuming that one connective is involved in relating each of these sentences. This is a lower percentage compared to the expected one in the Arabic corpus, where connectives relate more than 55% of sentences.

6. Connectives involved in medium range connectivity are higher in number in Arabic. The corpus contains 489 such connectives representing approximately 3% of connective mass. In comparison, the English corpus contains 195 such connectives, which correspond to 2% of total connective mass.

7. The influence of the frequencies mentioned in (6) above is better displayed when they are related to the number of paragraphs they operate within. It is estimated that medium range connectives function within approximately 5% of the paragraphs that constitute the corpus, assuming one operational occurrence within each of these paragraphs. In comparison, Arabic medium range connectives operate within 16% of paragraphs in the corpus.

8. Further, Arabic utilises more connectives to establish long range connectivity. There are 2,331 connectives in the corpus representing approximately 14% of total connectives. English, in contrast, uses 1,181 connectives that stand for 12% of the total number of connectives in the corpus.

9. The variation mentioned in (8) above is indicative of a stronger tendency in Arabic than English to relate paragraphs via the use of connectives. This is more clearly evident if we relate the absolute frequencies of connectives used for long range

connectivity to the number of paragraphs in each corpus. It is estimated that in English such connectives relate 29% of the paragraphs in the corpus, assuming one operational occurrence for each of these paragraphs. In contrast, Arabic long range connectives function across 77% of the paragraphs in the corpus, a distinctly higher percentage.

10. The organising force of connectives can be understood better if we investigate the connective range of operationality within each functional category. Different connective categories utilise range differently in each corpus and therefore each category reflects different tendencies towards text organisation. The exploration of this point is a task for the rest of this chapter.

9.3.4 Range of Additive Connectives

Variations in the distribution of range of additive connectives are displayed in Table (9.8) and are summarised below:

1. Generally, the distribution manifests similarity in the ordering of the types of range within both corpora. The most frequent type is immediate range, followed in descending order, by short, long and medium ranges.

2. However, there are variations in the absolute and relative frequencies of each range. The immediate range characterises additive connectives in English as it constitutes approximately 66% of the total frequency of the category. This indicates smaller frequencies for the rest of the categories. In contrast, Arabic additives, though they are still characterised by a high relative

frequency of immediate range, manifests higher frequencies for the other types of range.

	<u>English</u>		<u>Arabic</u>	
Range ----	Abs. ----	Rel ---	Abs. ----	Rel ---
Immediate	1487	65.80	4435	51.86
Short	449	19.87	2352	27.50
Medium	49	2.17	248	2.90
Long	275	12.16	1517	17.74
-----	-----	-----	-----	-----
Total	2260	100.00	8552	100.00

Table 9.8 Distribution of range of additive connectives

3. The distribution of range confirms that there is a stronger tendency in Arabic towards using additives for achieving a short range connectivity, i.e. connectivity of adjacent sentences. If we relate the absolute frequency of short range additive connectives to the total number of sentences in each corpus (11,671 in English and 8,060 in Arabic), we find that, on average, this type of range operates on 4% of sentences in English but 29% of sentences in Arabic. This difference is significant enough to characterise the operational mode of additive connectives in each language.

4. Furthermore, Arabic additive connectives have a stronger tendency for achieving a long range connectivity, i.e. relating paragraphs. If we compute the absolute frequency of long range additive connectives against the total number of paragraphs in each corpus (4,083 in English and 3,033 in Arabic) we find that, on average, additive connectives are responsible for combining less than 7% of the paragraphs in English and 50% in Arabic. Again the

difference is significant enough to be considered characteristic of each language.

5. The difference between the relative frequencies of medium range additive connectives is small (2.17% in English vs. 2.9% in Arabic). However, if the absolute frequencies are related to the total number of paragraphs, the difference is magnified. On average, the connectives are operational in relating sequences within 1.2% of the paragraphs in English and 8.2% in Arabic. This shows a stronger tendency in Arabic to resort to additive connectives to establish connectivity of this type of range.

9.3.5 Range of Comparative Connectives

The distribution of range of comparative connectives in both corpora is given in Table (9.9). The results are summarised below:

1. In both corpora the most frequent operational range is that of the immediate type followed by the short type. However, there is a difference in the distribution of the two other types. In English, long range is more frequent than medium range operationality. In Arabic, both types have a similar frequency.

2. In general, comparative connectives are characterised by an immediate operational range. This is evident in the high frequency of this type of range in both corpora. However, the tendency seems more prominent in Arabic, where comparative connectives with immediate range constitute 94% of the connectives in the category compared to 87% in English. This means slightly higher frequencies for the other types in English.

<u>Range</u> -----	<u>English</u>		<u>Arabic</u>	
	<u>Abs.</u> -----	<u>Rel</u> ---	<u>Abs.</u> -----	<u>Rel</u> ---
Immediate	383	86.85	251	94.36
Short	32	7.26	9	3.38
Medium	2	0.45	3	1.13
Long ----	24 ---	5.44 -----	3 ---	1.13 -----
Total	441	100.00	266	100.00

Table 9.9 Distribution of range of comparative connectives

3. Compared to their Arabic counterparts, English comparative connectives are characterised by a stronger tendency towards a short range intersentential connectivity. Not only connectives with this type of range have higher share in the total comparative mass (7.26% in English compared to 3.38% in Arabic), but they are operational in a higher percentage of sentences in the English corpus. They combine 0.27% of sentences in English while their Arabic counterparts operate on 0.11%.

4. Further, there is a stronger tendency in English than in Arabic to utilise comparative connectives for long range connectivity. Connectives having this type of range constitute 5.44% of comparative mass in English and 1.13% in Arabic. In English, long range comparative connectives operate on 0.6% of paragraphs in the corpus. Their Arabic counterparts operate on only 0.1%.

5. Only a small percentage of comparative connectives function

with a medium range of connectivity. These connectives constitute a slightly higher group in Arabic than English (1.13% of Arabic comparatives compared to 0.45%). This is indicative of the smaller role that English comparative connectives play in relating "unadjacent" segments within the paragraph boundaries.

9.3.6 Range of Alternative Connectives

Differences in the distribution of operational range of alternative connectives are displayed in Table (9.10) and summarised below:

1. In general, the distribution exhibits similarity in the ordering of the four types of range. In both corpora, the order by descending frequency is immediate, then short, long and medium ranges.

	<u>English</u>		<u>Arabic</u>	
<u>Range</u>	<u>Abs.</u>	<u>Rel.</u>	<u>Abs.</u>	<u>Rel.</u>
Immediate	121	84.61	162	82.65
Short	16	11.19	26	13.27
Medium	1	0.70	0	0.00
Long	5	3.50	8	4.08
Total	143	100.00	196	100.00

Table 9.10 Distribution of range of alternative connectives

2. Basically, alternative connectives in both corpora are characterised with an immediate range of operationality. This is evident in the high and closely similar relative frequencies of connectives having this range (85% in English and 83% in Arabic).

3. Alternative connectives having a short range constitute 11.19% of the connectives in the category in English and 13.27% in Arabic. However, the difference is magnified when we relate their absolute frequencies to the total number of sentences in each corpus. On average, English short range alternatives operate on 0.14% of the sentences in the corpus while their Arabic counterparts combine 0.32% of sentences.

4. Further, alternative connectives have a small role in relating paragraphs or larger stretches. Those that have a long operational range constitute 3.5% of connectives in the category in English and 4% in Arabic. Related to the number of paragraphs in each corpus, English alternatives combine 0.12% of paragraphs while their Arabic counterparts operate on 0.26%.

5. Alternatives have not been observed to have a medium range of operability in Arabic. In English, there are only one occurrence in the corpus. This reflects a tendency in both corpora not to employ alternatives for this type of organisation.

9.3.7 Range of Reformulatory Connectives

The distribution of range of reformulatory connectives is given in Table (9.11). The results are summarised below:

1. In general, the distribution exhibits similarity in the frequency ordering of the four types of range in both corpora. The order, by descending frequency, is short range, followed by immediate, long and medium ranges.

<u>Range</u> -----	<u>English</u>		<u>Arabic</u>	
	<u>Abs.</u> -----	<u>Rel.</u> -----	<u>Abs.</u> -----	<u>Rel.</u> -----
Immediate	33	28.69	78	32.64
Short	45	39.13	120	50.21
Medium	9	7.83	10	4.18
Long ----	28 --	24.35 -----	31 --	12.97 -----
Total	115	100.00	239	100.00

Table 9.11 Distribution of range of reformulatory connectives

2. But despite this similarity, there is a stronger tendency in Arabic to utilise reformulatory connectives for short range connectivity. These connectives represent more than 50% of total uses of the reformulatory category in Arabic, compared to 39% in English. When related to the total number of sentences in each corpus, these connectives operate on 1.5% of the sentences in Arabic compared to 0.4% in English.

3. There is also a stronger tendency in Arabic for using reformulatory connectives for immediate range connectivity. This is evident in the higher absolute and relative frequencies of such uses in Arabic compared to English.

4. Both corpora utilise reformulatory connectives for long range connectivity. The number of such uses, though higher in terms of absolute frequency in Arabic (31. vs. 28), is lower in terms of relative frequency, i.e. the share that such uses represent in the total reformulatory uses in the corpus (13% vs. 24% in English). However, when operationality is computed in the corpus, we have

found that Arabic reformulatory connectives relate more than 1% of paragraphs in the corpus while their English equivalents operate on less than 0.7%.

5. Reformulatory connectives that are utilised for medium range connectivity constitute approximately 8% of their total uses in English and are expected to operate within 0.22% of the paragraphs in the corpus. In Arabic, such connectives represent 4% of the uses of reformulatory connectives and are estimated to operate on 0.33% of the paragraphs in the corpus.

9.3.8 Range of Orientative Connectives

Variations in the distribution of orientative connectives are given in Table (9.12). Below is a summary of the results:

1. In general, the relative frequency of the four types of range manifests some similarities across the two corpora. Ordered by descending frequency, the most frequent type is that of immediate range followed by short, long and medium.

2. Although orientatives involved in immediate range connectivity are higher in English in terms of absolute frequency (481 vs. 377), they are lower in terms of relative frequency. Approximately 48% of the orientatives in English are used for this type of range compared to 51% in Arabic.

3. There is a relatively stronger tendency in English to use orientatives for short range connectivity, i.e. establishing links across sentences. This is evident in the higher absolute and

relative frequencies of orientatives utilised for this purpose. The computation of operationality across the sentences in the corpus shows that, on average, orientatives operate on 2.6% of sentences in English compared to 2% in Arabic.

	<u>English</u>		<u>Arabic</u>	
<u>Range</u>	<u>Abs.</u>	<u>Rel.</u>	<u>Abs.</u>	<u>Rel.</u>
Immediate	481	47.58	377	51.08
Short	303	29.97	167	22.63
Medium	34	3.36	22	2.98
Long	193	19.09	172	23.31
Total	1011	100.00	738	100.00

Table 9.12 Distribution of range of orientative connectives

4. Although there are more English orientatives operative across the paragraph boundaries (193 compared to 172 in Arabic), their uses constitute a lower size of share in the total utilisation of orientatives (19% compared to 23% in Arabic). These numbers mean that orientatives operate across 4.7% of paragraphs in English and 5.7% in Arabic.

5. Orientatives utilised for medium range connectivity constitute 3.36% of total orientatives in English and 3% in Arabic. The absolute frequencies (34 compared to 22 in Arabic) means that these connectives operate within 0.83% of the total paragraphs in English and 0.72% in Arabic.

9.3.9 Range of Temporal Connectives

Distribution of range of temporal connectives is given in Table (9.13). The main results of the comparison are summarised below:

1. The distribution across the two corpora indicate that English utilises a slightly larger share of temporal connectives for long range connectivity. The corpus includes 125 such connectives representing 9.74% of temporals. In comparison, Arabic exploits 114 connectives representing 7.63% of the category.

	<u>English</u>		<u>Arabic</u>	
<u>Range</u>	<u>Abs.</u>	<u>Rel.</u>	<u>Abs.</u>	<u>Rel.</u>
Immediate	990	77.10	1100	73.63
Short	140	10.90	256	17.13
Medium	29	2.26	24	1.61
<u>Long</u>	<u>125</u>	<u>9.74</u>	<u>114</u>	<u>7.63</u>
Total	1284	0.31	1494	100.00

Table 9.13 Distribution of range of temporal connectives

2. There is a stronger tendency in Arabic for using temporal connectives for short range connectivity. The corpus contains 256 such connectives, representing 17% of temporals. The equivalent number in English is 140 connectives representing 11% of temporals.

3. Although Arabic has a larger number of temporal connectives having an immediate operational range (1,100 in Arabic vs. 990 in English), the share of temporal mass that this number represents is slightly smaller (74% compared to 77%). In both corpora this range

is the most frequent in the corpus.

4. In English temporal connectives with a medium range of connectivity represents a higher share in English (2.26% compared to 1.61%). This indicates that English temporals play a more active role in organisation across the paragraph.

5. Four connectives in English have been identified as having indeterminate range. This is due mainly to their exophoric function in the text. These have been included with long range connectives.

9.3.10 Range of Spatial Connectives

Distribution of range of spatial connectives is given in Table (9.14). The variations are summarised below:

1. There is a stronger tendency in Arabic to utilise spatial connectives almost exclusively for immediate range connectivity. This is evident in the relative frequencies for this type of range in both corpora (95% in Arabic compared to 84% in English).

Range -----	<u>English</u>		<u>Arabic</u>	
	Abs. ----	Rel. ----	Abs. ----	Rel. ----
Immediate	42	84	38	95
Short	4	8	1	2.5
Medium	1	2	0	0.0
Long ----	3 --	6 -----	1 --	2.5 -----
Total	50	100.00	40	100.00

Table 9.14 Distribution of range of spatial connectives

2. Relative to Arabic, English spatial connectors are more frequently used to achieve short and long range connectivities. This is confirmed by the higher absolute and relative frequencies of these types of range in English in comparison to Arabic. This means that Arabic spatial connectors have a very small role in sustaining cohesion across sentences or larger text sequences.

9.3.11 Range of Causal Connectives

Distribution of range of causal connectors is given in Table 9.15 and the results are summarised below.

1. The distribution displays the reliance of causality on immediate range connectivity. Connectors having this function constitute more than 84% of causal connector mass. In comparison, Arabic utilises a relatively smaller share of causal connectors for immediate range of connectivity (59%), although in terms of absolute frequency the number is larger.

	<u>English</u>		<u>Arabic</u>	
<u>Range</u>	<u>Abs.</u>	<u>Rel.</u>	<u>Abs.</u>	<u>Rel.</u>
Immediate	1593	84.37	2246	59.45
Short	177	9.38	1082	28.64
Medium	20	1.06	145	3.84
Long	98	5.19	305	8.07
Total	1888	100.00	3778	100.00

Table 9.15 Distribution of range of causal connectors

2. There is a stronger tendency in Arabic to utilise connectives for short range connectivity. There are 1,082 connectives having this range compared to 177 in English. In terms of relative frequency, Arabic employs approximately 29% of causals for establishing this type of operational range, compared to 9% in English. When the numbers are related to the total number of sentences in the corpora, causal connectives in Arabic are estimated to operate on 13.4% of sentences while their equivalents in English are expected to operate on 1.5%. This important finding indicates that Arabic is more prone than English for explicit signalling of causality when the relation obtains across adjacent sentences.

3. Another characteristic of range is the strong tendency in Arabic towards using causal connectives for long range operability. There are 305 connectives in Arabic involved in this type of range, which represents 8% of causals. In contrast, English employs 98 connectives, corresponding to 5% of its causal connective repertoire, to achieve across the paragraph connectivity. Related to the number of paragraphs in each corpus, causal connectives operate on 10% of the paragraphs in Arabic and 2.4% in English.

4. Furthermore, Arabic contains more connectives utilised to achieve medium range connectivity. The corpus includes 145 such connectives, representing approximately 4% of the category. In comparison, there are 20 connectives in the English corpus involved in medium range causality, a figure that corresponds to only 1% of causal connectives.

9.3.12 Range of Adversative Connectives

Distribution of range of adversative connectives is given in Table (9.16). The variations are summarised below:

1. There are distinct variations in the size of utilisation of each type of range across the two languages. Not only there are differences in the absolute frequency of connectives having a particular operational range, but there are also differences in their relative frequencies. However, as with most categories the order of the frequency of the four types of range is similar in both corpora. Sorted by descending frequency the order starts with immediate range, the most frequent, then short, long and medium ranges.

Range -----	<u>English</u>		<u>Arabic</u>	
	Abs. -----	Rel. -----	Abs. -----	Rel. -----
Immediate	1205	50.12	1030	60.87
Short	719	29.91	445	26.30
Medium	50	2.08	37	2.19
Long ----	430	17.89	180	10.64
Total	2404	100.00	1692	100.00

Table 9.16 Distribution of range of adversative connectives

2. There is a stronger tendency in Arabic to utilise adversative connectives for achieving immediate range connectivity. This is confirmed by the relative frequency of this type of range across the two languages. In Arabic, 61% of connectives are used for this

purpose compared to 50% in English.

3. In contrast there is a stronger tendency in English to utilise adversatives for achieving a short range connectivity. The corpus comprises 719 adversative connectives, representing 305 of the category, that have this type of range. Related to the number of sentences in the corpus, they represent operationality within 6.2% of sentences in the corpus. Arabic, on the other hand, utilises adversatives for this type of range to a lesser degree than English. The total number of adversatives having a short operational range is 445 representing 26% of the category. Related to the number of sentences in the Arabic corpus, they are estimated to operate within 5.5% of sentences in the corpus.

4. Furthermore, English has a stronger tendency towards utilising adversative connectives for achieving long range connectivity. The number of adversatives having this type of range is 430, representing 18% of this category in the corpus. These are operational within an average of 10.5% of paragraphs in the corpus. In comparison, Arabic comprises a smaller set of adversatives utilised for long range connectivity. The set comprises 180 representing 10.6% of the category in the corpus. This set is operational within approximately 6% of the number of paragraphs in the corpus.

5. Both corpora utilise a closely similar shares of adversatives for achieving medium range connectivity. These shares, however, are small compared to those reserved for other types of range.

9.4 Conclusion

The main aim of this chapter has been a contrastive discussion of the textual variations, both quantitative and qualitative in the behaviour of connectives in the English and Arabic corpora. This primary aim has been achieved by first examining some statistical features of connectives across the two corpora, including their frequency distribution, connective-sentence relation, patterns of repetitiveness, growth, entropy and redundancy, and a host of other features such as concentration, extensity, exclusivity and density. This detailed examination displays distinct variations that are language specific and therefore represent characteristics of connectivity in each language.

Comparison is then extended to a consideration of each functional category of connectives. The procedure adopted is summarised as follows. Each category is examined individually across the two languages. Comparison starts by bringing in the quantitative accounts of the category in each language as introduced in Chapter 8. Observations and comments are made on the shared and unique patterns of frequency, repetitiveness and growth. Each contrastive account is supported by a visual display of the variations. Comparison is also made of the main features of the subcategories within each category: frequency, repetitiveness and the behaviour of the top most frequent connectives.

This comprehensive quantitative comparison is then followed by a detailed discussion of the textual variations in the functioning and patterning of each subcategory. Extensive exemplification is used for illustrative purposes. The general conclusion is that

connectives within each category manifest considerable variations in the way they combine sequences to create text. The variations do not only reside in the imposition of a specific surface organisation on the connected objects, but they often extend to include conceptual formatting of the knowledge configurations involved in the connection. In other words, connectives behave differently in each language in the way they uphold sequential and conceptual connectivities and therefore they assist in achieving different text linearisation.

Another aspect of the textual behaviour of connectives that is examined across the two languages is the range of operationality of connectives. Four types of range have been identified. These are labelled for convenience: "immediate", "short", "medium" and "long". The organisational force of each type is first discussed in general terms. Later a detailed discussion is made for ranges in each corpus. It has been noticed that Arabic manifests stronger tendencies for utilising connectives to establish short, medium and long range connectives.

This discussion is supported further by an exposition of range within each of the nine main categories of connectives. One similar feature is that the frequency ordering of the four types of range within most categories is the same. However, when examined on a one-to-one basis, each type manifests variations that reflect inherent possibilities or limitations for text organisation. The variations take the form of tendencies that characterise the connectives of a particular category in one language as opposed to

the other and are therefore language specific.

What is required next is to draw some generalisations from the detailed discussion that we have presented both in this and previous chapters. These generalisations are to be intended to offer some guidelines for the design of instructional material for EFL teaching of writing. This task is left to the next, final, chapter.

Footnotes to Chapter 9

(1) This again exemplifies the general fuzziness of compartmentalising textual relationships. One important underlying factor is the degree and complexity of semantic blend that exists in the expression of relations. However, I would argue that this factor differs qualitatively and quantitatively across the two languages.

(2) This pattern has been monitored in the corpus with the assistance of the computer (see Ch. 5). A SPITBOL program is written that will strip the corpus of all word occurrences except connectives. The result is a map where the role of connectives in paragraph building can be closely inspected.

(3) Some classical grammarians believe that "'idā" is fundamentally temporal and that its conditional connotation is "not familiar" (Hasan 1980 Vol.4 p.440 fn.2). Others believe that "'idā" may lose its conditional connotation only in certain cases, such as the absence of "fa" from statement B, i.e. the consequence (ibid p.441).

CHAPTER 10

Conclusions and Pedagogical Implications

10.0 Perspective

This final chapter intends to achieve three related tasks. First it reconsiders the evidence that has emerged within the description of the textual functioning of connectives in English and Arabic and in the subsequent contrastive analysis. The aim is to arrive at generalisations concerning the predominant types of textual patterning that are characteristic of text organisation in each language. This involves assessment of types of connectivities and their role in influencing operations of text linearisation.

The second task is applicational. Findings arrived at in the first task are incorporated in a scheme that suggests pedagogical guidelines for designing a learner-centred type of written composition practice.

The third task lists a number of proposals for future extension of this project. These can be grouped in three areas of inquiry: one is related to interlingual investigation in text structure; the other concerns the utilisation of the corpora for computer-assisted research; the third involves a number of lexicographical projects; and the fourth involves further research in EFL teaching of written text composition.

Accordingly, the plan of the chapter comprises four sections. The first considers aspects of text organisation in English and Arabic, particularly those that involve use of connectives. The

second considers types of connectivities and attempts at distinguishing types of strategies and textual choices favourable in each of the two languages. The third section examines some aspects of the EFL teaching operation and relates to it the findings in the first two sections. The last section makes suggestions for further research.

10.1 Text Organisation

Text as a written manifestation of language is organised in sequences that usually reflect the underlying conceptual organisation of events, actions, situations and objects in the textual world. These sequences differ in size and complexity and in the way they are linearised within the text or across various texts. Linearity entails a hierarchy of grouping of various text elements into larger "chunks", and grouping suggests methods by which the various elements and segments are sequenced, i.e. related to each other. This, in turn, means that certain parts of the text are perceived as more salient than others. Linearity is thus regulated by activities of textual grouping, sequencing and achieving salience. (Refer to the discussion of text linearisation in Ch. 3)

Grouping and hierarchisation exist on syntactic, orthographic and rhetorical levels. Normally there is a great degree of compatibility in textual segmentation and grouping at these three levels. This is particularly true when grouping involves the creation of a major unit of written text: the sentence. On the syntactic level, a sentence, apart from some areas of indeterminacy, may consist of one clause, making up what is traditionally called

the simple sentence, or more than one clause, i.e. a clause-complex. On the graphological level, the sentence in English starts with a capital letter and ends (in English and Arabic) with a terminator such as a full-stop (but cf. the discussion below), a question or exclamation mark. On the conceptual level, the sentence may reflect the informational content of one or more propositions.

Grouping of sentences produces the next textual unit in the hierarchy: the paragraph.⁽¹⁾ This unit is delineated on the orthographic, syntactic, and rhetorical/conceptual levels. Orthographically, it is clearly marked with indentation, though this type of visual marking exhibits some degree of indeterminacy (see arguments in Longacre 1979).⁽²⁾ As a syntax-based format, the paragraph may consist of as many clauses or clause-complexes as befits the principle of "salience" (see Chapter 3). Rhetorically, the paragraph is regarded as a unit that maintains a certain unified orientation (Grimes 1975, Hinds 1979), the most notable being the thematic orientation. This means a continuous theme is represented in the unity of a text partition; that is, all components (sentences) contribute to the realisation and local expansion of a topical locus. Once a thematic discontinuity takes place, a demarcation boundary is made to terminate the partition.⁽³⁾

Grouping may also produce units that are intermediary between the paragraph and text (when the latter is sizeable enough to comprise a number of paragraphs) and various labels have been suggested (such as the "cluster", see Chapter 3). The basis of this hierarchical level of grouping is rhetorical/conceptual: the unit reflects

closely related stages in the development of a particular topic.

Text linearity is further regulated by a set of operations of sequencing, comprising activities and procedures that aim to arrange the various components in a textual order. Operations of sequencing (refer to discussion in Chapter 3) regulate text organisation via activities of juxtaposition, regression, progression and pausing. These are text-creating operations that reflect themselves in sequential and conceptual connectivities. In text actualisation these types of connectivities organise the various sequences in a usable format and their efficient utilisation secures efficiency of text design.

Another regulative factor that affects text organisation concerns the achievement of salience. This pertains to textual activities that are aimed towards producing intense intrusions upon sensory apperception. The activities affect choices made at various stages of text actualisation and are therefore congruent with both the text plan and the text producer's goals. The efficient utilisation of these activities gives text its textual effectiveness.

10.2 Text Organisation in English and Arabic

10.2.1 Preliminaries

Even a cursory inspection of the numerical analysis of textual organisation in English and Arabic (Chapter 7) will confirm three general conclusions:

a) Tendencies for textual grouping differ across the two languages and this is manifested in the considerable variation in

the size and complexity of the textual sequences that constitute text in each language.

b) Textual sequencing differs in type and extent of operationality and this is reflected in the interlingual differences in textual connectivity.

c) Most importantly, there are significant variations in the central roles that connectives play in text construction.

We shall undertake within this section to discuss the variations in points (a) and (b) within a broad textual perspective, drawing upon the results of the discussion in Chapters 3 and 4 and the analysis in Chapters 7 and 9. Point (c) will be considered in the next section.

10.2.2 Textual Grouping in English and Arabic

One distinctive variation in textual grouping of constituents in English and Arabic is the size of the resulting text "chunk". Arabic tends to group constituents into segments of larger size than English does. Following the description of the concept of word as a unit of linguistic measurement in Appendix (1), and drawing on the results of the calculation of sentence length given in Chapter (7) and Appendices (10) and (11), we conclude that sentence length is larger in Arabic in comparison to English. The average sentence length is computed as 22 words in English and 32 in Arabic. This, in effect, means that Arabic tends to segment the text into fewer, but larger, sentences than English does and is exemplified in the smaller number of sentences that make up the Arabic corpus: 8,060

compared to 11,671 in English.

A related distinctive difference of textual grouping concerns the variability of sentence length within each corpus. Arabic exhibits a larger degree of variability than English. Sentence length in the English corpus varies from 1 to 103 words, while sentence length in Arabic ranges between 1 and 212 words. Computation of the degree of variability gives a coefficient of 55% in English and 90% in Arabic.

The next segment in the textual hierarchy, i.e. the paragraph, is, like the sentence, of a larger size in Arabic compared to English. The average paragraph length is approximately 63 words in English and 85 words in Arabic. This means that Arabic segments the text into fewer but significantly larger paragraphs than English does and is confirmed by the smaller number of paragraphs in Arabic: 3,033 compared to 4,083 in English.

A related feature that is characteristic of Arabic is the larger degree of variability of paragraph size compared to English. Paragraph lengths vary from 1 to 250 words in English while in Arabic it ranges from 1 to 1,306 words. In general, the extent of the variability is stipulated in the coefficient of variation, which is calculated as 44% in English and 85% in Arabic.

10.2.3 Textual Sequencing: The Nature of Connectivity

Textual sequencing involves a variety of activities that impose a certain linear format to the underlying relational configurations. In Chapter 3 we have distinguished four major mapping operations that influence text organisation from the phases of planning and

ideation all the way to the linearisation of the surface text. These operations are juxtaposition, regression, progression and pausing: they represent the way textual cohesion as a process is initiated and sustained.

One of the significant means of achieving textual sequencing and so regulate effectively both text organisation and conceptual development is the use of connectives (see discussion in Chapter 4). These devices are involved in procedures that aim to facilitate the conduct of sequencing operations and, as a result, render the utilisation of text feasible under normal conditions. It is our contention that variations in the textual distribution, functioning and patterning of connectives, when taken collectively, contribute to a large extent to variations in overall cohesion and text organisation. This has been confirmed by the empirical evidence we have gathered from the computer-aided analysis of connectives that has been reported in this thesis.

Results of the analysis point to the following findings:

1. A major bi-lingual difference in text organisation is the heavier reliance of textual sequencing in Arabic on the explicit utilisation of connectives in comparison to English. This is made visible by the results of several calculations:

- a. There is a highly significant difference in the frequency distribution of connectives in the English and Arabic corpora. The Arabic corpus contains a higher number of connectives (16,995) than English (9,596).

b. The analysis of the frequency of connective per sentence distribution confirms the significant role of connectives in Arabic in relating text constituents in comparison to English (see discussion in 9.1.2). For instance sentences that do not contain any connectives, and therefore relies for cohesion on the exploitation of other types of cohesive means, constitute 46% of total sentences in English, but 11.6% in Arabic.

c. The analysis of connective range (cf. 9.3.3) has shown that there are stronger tendencies in Arabic to use connectives to relate sentences within a paragraph and paragraphs within a text. These tendencies are exhibited in the higher absolute and relative frequencies of connectives of short, medium and long ranges of operationality.

d. The analysis of interval distribution has exhibited that one characteristic feature of Arabic connectivity is the short distance that separates occurrences of connectives (see Chapter 7). The average distance is 14 words in Arabic compared to 26 in English.

e. The analysis of repeat rate is an evidence for the significant role of connectives in preserving cohesion. One of the results of the analysis (see also point 5 below) shows that the probability that two words occurring in succession in the corpus will turn out to be connectives is significantly higher in Arabic than in English (44×10^{-4} vs. 14×10^{-4}) (see relevant analysis in Chapters 7 and 9).

2. Another difference that is related to point (1) above is the

complexity of the sequences. The average number of connectives per sentence in English is 0.82 and in Arabic 2.1. These figures are indicative of a strong tendency in Arabic to organise a big number of clauses in one long sequence. This method of text organisation is responsible for the overt inconsistency of the use of punctuation, particularly the use of the comma and full-stop. The text producer simply assumes that the sequence is still continuous, and would therefore like to relate to it as many compatible sequences as textual efficiency permits, without the need to break the sequence with a pause. In other words, a number of conceptual configurations are related one to the other, particularly via the use of connectives, and the process is prolonged until the conceptual "chunk" is exhausted, i.e. totally realised and incorporated in the surface format, and a point of termination becomes necessary. For instance, at the end of a paragraph a point of termination is necessary and the paragraph is therefore clearly marked; but, there is an apparent lack of consistency in allocating a point of termination to a sequence of closely knit propositions within the paragraph. This is encouraged by the fact that Arabic textuality tolerates long sequences to an extent that is not permissible in English.

3. The heavy reliance on connectives for organising text and sustaining cohesion in Arabic stems from a textual pressure to preserve continuity and integration of propositional development. Arabic is extremely sensitive to discontinuity, that is a sudden disintegration that is sufficiently prominent as to create a gap that disturbs text stability. This gap is immediately bridged with

a connective, usually an additive: "wa" or "fa".

There are several organisational factors that create discontinuity in Arabic, the most frequent are two:

a. A number of connectives when used in initial position in the sequence create a discontinuity and have therefore to be supported by an additive connective to establish the missing link.

b. Often adverbials, including those operating as connectives, create discontinuity when used in a thematic position in the sequence followed by a nominal clause. To bridge the gap, "fa" is used to integrate the adverbial with the subsequent stretch of text. In addition "wa" may occasionally, particularly in the absence of other supportive connectives, link the whole of the sequence with the previous one.

In contrast, these factors have not been observed to create similar disintegration in English. Text cohesion in English relies on a variety of hard-core and soft-core devices to create, relate and organise text sequences (see discussion in Chapter 3). Excessive use of connectives, while tolerated and often commendable in Arabic, is normally a sign of inefficient textual design in English. That is, English connectives contribute to efficiency of design as long as their use is not unduly frequent. The textual strategy that is often adopted is to use connectives to signal relations that are not readily inferable, either because they are variable or non-expected. Connectives are also used where signalling relations is responsive to the demands of effectiveness or appropriateness of text design (see Chapter 3).

4. The factors mentioned in (3) above (in addition to others that will be discussed in the next section) explains why Arabic relies on the utilisation of a small number of connective types to achieve concatenation of textual sequences. The connectives "wa" and "fa", the first two most frequent connectives, have a combined frequency equivalent to more than 59% of connective tokens in the corpus. In English, this size of token coverage is achieved by the top ten connectives.

5. Related to (4) above is the tendency of text organisation in English to exploit more varied connective types for signalling various relations and concatenating text components. In other words, text organisation in English displays richer and more variable use of connectives in comparison to Arabic. This is reflected in the TTR in both corpora calculated as 0.032 in English and 0.017 in Arabic. It is also reflected in the higher predictability index of Arabic connectives in comparison to English (cf. calculations in Chapters 7 and 9).

The differences rest on a distinction between intensity and extensity of utilisation of connective types for sustaining cohesion. Arabic tends to make an intensive use of a relatively limited set of types while English tends to achieve an extensive exploitation of its available repertoire. This is confirmed by a variety of indices reported in Chapter 7. Each index is considered a characteristic feature of connectives in the relevant language. For instance, English connectives exhibit higher hapax probability, higher exclusivity index (representing the size of the share that

hapaxes reserve in the token mass), higher variegation index (indicative of connective diversification), higher type occurrence rate, higher gravity index (representing the occurrence rate of hapaxes) and higher concentration.

In contrast, Arabic connectives exhibit higher indices of repeatedness: higher token occurrence rate, higher non-hapax probability, higher index of stereotypicality (referring to the repeatedness of non-hapaxes), higher index of rhythmicity (indicative of the extent of repeatedness), and higher consolidation factor.

Another evidence that supports the tendencies towards intensive vs. extensive utilisation of types is the results of the calculation of repeat rates. For instance, the probability that two successive words in the corpus will turn out to be the same connective is significantly higher in Arabic than English (1.093×10^{-3} vs. 0.088×10^{-3}). Further, the probability that two successive connectives, with or without intervening gap, will be repetitions of one or the other connective is significantly higher in Arabic than in English (0.25 vs. 0.06) (refer to the detailed description in Chapters 7 and 9).

6. A further aspect of the intensity vs. the extensity of utilisation of connectives is the relation between the number of types and length of text, which has been investigated in the analysis of growth (see the discussion in Chapters 7 and 9). In examining the growth of connective types within successive sections of equal size in each corpus, it has been noted that there is a

faster saturation of types in the English than in the Arabic corpus. This is confirmed by analysing the growth of connective types within text intervals containing equal number of connective tokens. On the other hand, inspection of growth of tokens within intervals of equal size in each corpus points to a significantly more intensive use of connectives in each interval in the Arabic in comparison to the English corpus.

7. Another aspect of the intensity vs. extensity of use of connectives in text organisation is related to the computation of entropy and redundancy of connectives (refer to the analysis in Chapters 7 and 9). Arabic connectives have a lower entropy (H) than their English counterparts (3.9859 vs. 5.4379). Entropy, as has been mentioned in Chapter 7, represents the number of binary choices required for guessing or identifying a particular symbol, in our case a connective. In Arabic, the probable number of guesses needed for restoring a missing connective is 4; in English it is 6. The difference is explained by reference to the higher repetitiveness of connective types in Arabic in comparison to their variability in English.

The measure of redundancy of connectives has been found to be higher in Arabic in comparison to English (71.64% vs. 58.89%). This indicates a greater degree of structure in the utilisation of connectives in Arabic. In addition, the high measure of redundancy is proportionate with a strong property that permits one to reconstitute a missing connective from a stretch of text with a reasonable degree of accuracy.

10.3 Text Organisation and Connectivity

10.3.1 Preliminaries

In categorising connectives we have grouped their functions into nine distinctive classes, each with its own functional domain and textual patterning. In this section we would like to reconsider these categories in the light of the empirical evidence that we have gathered and analysed. The aim is to proffer a clearer image of the textual functioning of connectives and enhance their organisational role in the text.

The motivation behind this account is fundamentally contrastive. We would like to capitalise on the analysis initiated in Ch. 6 and developed in later chapters in pinpointing the favourable options and prominent directions in connectivity and the type of rhetorical development they impose on the text in each language.

10.3.2 Types of Connectivity

We would like to propose that, textually, connectives are involved in achieving four types of connectivities. These represent strategies for text organisation and reflect the ways in which the configuration of concepts are activated and related in regard to a text. They are capable, for instance, of ordering, developing and/or relating a) the structures of events, actions, objects and situations; b) the expression of views, claims, arguments and reactions, c) the interpretation of human experience (apperception, emotion, etc.); and d) the setting of communicative/rhetorical contingencies of various types (e.g. equivalence or opposition).

Furthermore, they are capable of operating within the boundaries of the sentence (relating two clauses), the paragraph (relating various sentences to the topical one), or across the paragraphs (relating parts of the text and achieving topical unity).

The four strategies, it is necessary to note, may be utilised without involving connectives. It often happens that events or actions are organised in temporal or causal sequences where the organisational relationship is clearly felt to be present but is not explicitly expressed. However, we would like to argue that occasionally the relationships expressed in this way differ in the precision with which they formulate textual sequences, and therefore any attempt to include them in the analysis leads to considerable indeterminacy. Moreover, the presence or absence of an explicit signalling of organisation via the use of connectives is one of the principal variables that is utilised for contrastive discussion.

The four connectivities are here designated: 1) Extension, 2) Variation, 3) Elaboration and 4) Enhancement (terms borrowed from Halliday 1985 but given a more detailed analysis). In discussing each, we shall consider its operational scope, type of activities involved and its role within both corpora.

10.3.3 Extension

10.3.3.1 Operational Scope

This type of connectivity represents the textual strategy of developing a text sequence by adding a new sequence that extends it. Basically the operation involves adding one block of knowledge to

the current one and thus achieving an organisation that relies on cumulation of sequences. For instance, if the current block is X, then the organisation will take the form of "X + Y + Z". The connective acts as an explicit organisational marker for grouping the sequences and achieving connection. As a textual strategy, this connectivity imposes a type of linearisation dominated by operations of progression. However, this is probably an overstatement, or an oversimplification, for, as we shall mention below, operations of regression are also involved, though, usually, to a lesser extent.

10.3.3.2 Types of Extension

Connectivity via extension is achieved by two types of strategies: addition and comparison. The textual role of connectives in signalling each of these strategies has extensively been documented in Chapters 6, 8 and 9. We shall, therefore, draw upon that description and concentrate, by way of conclusion, on outlining the extent to which each set of textual activities assist in organising text in each language.

10.3.3.3 Extension via Additivity

10.3.3.3.1. Textual Role of Additivity

Additivity is considered a fundamental, yet powerful, strategem for organising text sequences. It has the effect of adding one block of text-world knowledge to another, forming a chain of individually identified conceptual configurations. It is, moreover, capable of operating at all levels of textual constituency: it relates clauses, sentences, paragraphs and longer sequences to each

other.

Extension via additivity requires the imposition of sameness of conceptual domains which the knowledge configurations are drawn from: two events, situations, states, etc. However, this requirement, particularly in Arabic, is often overlooked and additive connectives are capable of combining unusual items drawn from ad hoc domains.

10.3.3.3.2. Additivity in English and Arabic

1. Additivity is the major and most powerful means of text creation in Arabic. This is evident in the high frequency and large size of repetitiveness of additive connectives in the Arabic relative to the English corpus (see Chapters 8 and 9 for details). The mean occurrence of additive connectives is one per sentence, a frequency that is not matched by any functional category in English.

It follows that in Arabic signalling relations and grouping sequences via the utilisation of additive connectives should be considered a default procedure, i.e. a standard procedure resorted to, or assumed to be stipulated, in the absence of other types of connective utilisation. Alternatively, it may be treated as "preference" procedure, i.e. a standard operation routinely favoured and selected over others when various, competing alternatives are open. In either case, the procedure manifests high predictability which is characteristic of textual organisation.

In contrast, the utilisation of additive connectives in English cannot be considered a "preference" nor a "default" procedure,

unless, of course, additivity is discussed in general terms and without particular recourse to the explicit use of connectives. Such a case is not our concern in this study and the issue of implicit additivity will be left for future research.

To understand the interlingual variations in the cohesive role of additivity, we need to look into the role of the functional subcategories that collectively constitute the main function of additivity.

1. Appending

a. One of the most prevalent strategies in text organisation in Arabic is the imposition of an "ordering" or "listing" to the various knowledge blocks in the text world and then append their various surface realisations via a set of additive connectives. The use of the connective facilitates the transitional points between one sequence and another across the text. Their absence is sufficient to create gaps within the text and may damage text cohesion and burden or even halt processing.

From a rhetorical perspective, text development in Arabic is maintained to a greater extent than English by expansion based on a closely appended chain of relevant statements. Typically, a topical statement or sequence is expanded by adding to it, normally via the use of additive connectives, as many relevant statements (expressing views, ideas, suggestions, facts, events or courses of actions) as the text producer thinks appropriate without overloading the sequence. The threshold of overload is, however, far more flexible in Arabic and therefore shows a considerable degree of tolerance

towards excessive appending. In this way, statements are tied up in a neat package where the topical sequence is developed via an exhaustive succession of relevant possibilities. The package reflects a single complex awareness of the patterns of experience.

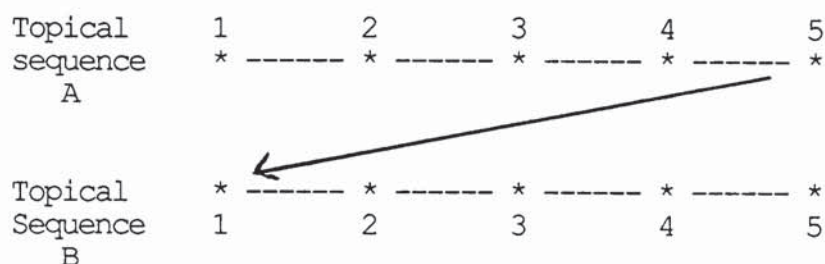


Fig. 10.1 Text development via appending in Arabic

Once the development of a sequence reaches a threshold of termination (which may coincide with the end of a paragraph) and another one is initiated, an additive connective of appending is more often used in Arabic for combining the two sequences into a larger one. This may happen even if the second sequence expresses topicality that is not directly related to the first one (see Figure 10.1).

b. Related to the discussion in (a) above is the tendency, stronger in Arabic than English, towards reinforcing the signalling of additivity with more than one connective (see Chapter 9 for an outline of the strong-weak additive connective distinction). This strategem is intended to perform a two-fold task: i) to strengthen an additive relation that would otherwise have been rhetorically damaged via the use of a "weak" additive connective; and ii) to reduce or even eliminate possible conflation of additivity/extension

with other contextually or rhetorically undesired activities, such as temporality.

c. One characteristic strategem that is far more often employed in Arabic than English is the use of parallelism to produce a rhetorical effect, for instance achieving persuasion. Parallelistic sequences are often combined additively, the most common connective being "wa", though occasionally asyndetic parallelistic constructions are also used.

Parallelism involves re-using of a surface format with different components. This is achieved by packaging the content of the subsequent proposition (or sequence) into already activated structural constructions, thus reducing processing load and leaving resources free to focus on content (see Chapter 3). Parallelism is maintained via complex activities of regression and progression that are combined to produce points of salience.

In Arabic, the degree of parallelism varies. Examination of this phenomenon in the corpus confirms what I suggested in Al-Jubouri (1984). In general, two degrees of parallelism can be distinguished with ample possibilities for variation.

A. Complete parallelism, where a whole surface construction is re-used, i.e. there is a total coincidence of the surface formats of text sequences. This can be combined, for greater effect, with one or both of the following schemes:

i. Use of similar rhyme to end each parallelistic sequence. This is referred to in Arabic rhetoric as "saj<", a classical scheme that

has appeared and developed into elaborate patterns in classical written Arabic, then started to decline when its rhetorical effect diminished through excessive use, particularly during "Middle" Arabic. The scheme is most effective when a) it is used sparsely within parallelistic forms, and b) the rhyming words are selected in such a way that they reflect natural development of the content of the sequence rather than appear as if imposed on it. Imposition of rhyming words often damages the efficiency of the scheme and impairs, sometimes drastically, its effectiveness.

ii. Isomorphism of content. That is, the conceptual content of the subsequent parallelistic sequence is a repetition, in different phrasing, of that of the antecedent. This is the essence of the functional additive relation of "coupling" (see also point 2 below), though coupling need not be identified with parallelism. The scheme is powerful and, when appropriately utilised, produces a massive rhetorical impact. The effect is achieved by combining two types of repetitions. First, there is a structural repetition which runs horizontally in the sequences and which is achieved by surface parallelism; that is, the subsequent sequence mirrors the structural organisation of the antecedent. The second type of repetition is realised vertically: the content of the antecedent and subsequent(s) constitute shared conceptual configuration (cf. Chapters 6 and 9, see also Al-Jubouri 1984 p.108).

The combination of schemes (a) and (b) results in a forceful type of parallelism that helps promote rapport between text producer and receiver. This type of parallelism is recurrent in Arabic expository discourse and can have a high persuasive effect.

B. Incomplete parallelism: This type exhibits a partial coincidence between the surface formats of text sequences. In other words, the parallelistic sequences have a similar, but not totally identical structural construction. Observation of this type of parallelism in the corpus confirms two points:

i. Incomplete parallelism occurs with more frequency than complete parallelism. This is due mainly to a rhetorical and/or conceptual need for enlarging or compressing the surface format of one or more parallelistic sequences.

ii. It is possible to combine more sequences in incomplete than in complete parallelism. This produces a chain of flexible parallelistic constructions which can be utilised to produce a cumulation of claims that renders an argument more persuasive.

2. Coupling

A related additive activity of extension the presence of which in Arabic constitutes a distinctive variation is the use of "coupling" (see discussion in 6.4.8 and 9.2.2.2.6). This is a reiteration of the same or similar conceptual configuration in additively connected propositions. The main textual motivation for the use of coupling is the establishment of salience via the creation of a loop whereby the processor accesses the same, or a closely related, semantic network more than once.

Coupling is part of the more general cohesive mechanism of repetition that is operative to a variable degree of intensity in

Arabic argumentative discourse (see Al-Jubouri 1984 for an outline of the discoursal role of repetition in Arabic). This mechanism functions on different levels of linguistic complexity. At the morphological level, repetition is realised in words that lie in close syntactic proximity, and is manifested in their root or pattern similarity (ibid pp. 100-102). At the word level repetition is realised through the use of what I call "word strings" (ibid p.102 ff), a structure that combines two or more words of identical syntactic function, i.e. nouns, verbs or adjectives, normally combined by an additive conjunction. At a text level, parallelism, coupling and paraphrase are some major complex manifestations of repetition.

Coupling, as inspection of the corpus confirms, is of two types:

a. coupling may be associated with parallelism (see discussion in 4 above). Here the connective, usually "wa", combines two propositions that represent structural and conceptual isomorphism.

b. coupling may be associated with paraphrase. In this type the conceptual content of the antecedent and subsequent remains largely constant while surface expressions are varied. Often the subsequent takes the form of a restatement of the antecedent. Rhetorically, this is part of a larger mechanism where arguments are restated a number of times in order to create the desired impact (see a brief discussion in Al-Jubouri 1984). The overall aim is enrichment of context and variation of perspective through which a view, argument, fact, etc., is stated.

3. Comment

Related to the previous two additive activities is that of "comment". This label subsumes not one but a group of related activities all having one essential textual role. They allow the text producer to pause and make a comment or express a viewpoint regarding the content of the current text sequence (the antecedent) (see discussion in 6.4.6).

A shared organisational role that "comment" has in both corpora is where the subsequent proposition (or set of propositions) enriches the conceptual content of the antecedent. This is performed by a) offering a definition of a vague or general configuration; b) expressing a clarification in order to disambiguate, that is in order to reduce or preclude unintended multiplicity of meaning; c) giving an explanation of a conceptual configuration that is not vague but that requires some elaboration in order, for instance, to rectify any deficiency, and d) making an evaluation related to the content of the antecedent.

But despite the presence of additive activities of comment in both corpora, it is significantly more prevalent in Arabic than English. An Arab writer is under constant pressure to define, clarify or evaluate the view, fact or course of action, that he has stated. Comment may sometimes centre on a particular concept expressed in the antecedent or may, alternatively, concern the whole conceptual content. In either case, the activities are related to and can be considered expansion of, those of paraphrase (see 4 above). Connectives involved in these activities constitute a small

set: "which", "and" and "after all" in English and "wa", "fa", "mimmā" and "alā wa".

Another organisational role that is unique to Arabic and that is highly frequent is the use of comment to establish focus (see 6.4.6.2 and 9.2.2.2.4). Text organisation is influenced by two types of operations. The first concerns bridging possible discontinuity and shifting focus to the subsequent. Arabic, as argued in 10.2.3 above, is highly susceptible to discontinuity, that is temporary disintegration in the flow of the text that can halt or impede processing activities. One main discontinuity occurs when a nominal sentence is introduced by an adverb or adverbial phrases, including those functioning as connectives. Continuity is restored via the use of "fa" which, in addition, places the conceptual content of the subsequent under focus.

The second type of operation results in intensifying the semantic space of certain expressions most of which are connectives (see examples in 9.2.2.2.4). The connective that performs this operation is "wa" which follows the expression it intensifies and precedes the subsequent. In this intervening position, "wa" brings the whole conceptual content of the subsequent under focus.

4. Enumeration

Another observed interlingual difference in the cohesive/organisational role of additive activities of textual extension rests on variation in modes of textual enumeration (see the detailed discussion in 6.4.4 and 9.2.2.2.2). In general, connectives of

enumeration play a vital role in text organisation. They dissect and order text segments into relatively discrete items that are distinctly comparable among themselves. On the conceptual level, the items should have a clear content relation that constitutes an immediately recognisable conceptual unity, so that processing can adapt accordingly.

The order in which the list is organised can be neutral. In this case the items are commutative and replaceable. Alternatively, the order can be made according to certain rhetorical factors such as scaling. This factor grades the items according to specific and predetermined degrees of prominence.

The interlingual variation observed reflects tendencies towards favouring one type of plan as opposed to another. In English there is a strong tendency towards using a plan with a broad system of itemisation. The plan normally has an initial, medial and final stages clearly marked by the connectives "first", "later", "then" and "finally". Arabic, on the other hand, favours a tight plan with discretely itemised ordering. This is signalled by resorting to numerals: "'awwalan", [first], "ṭāniyan" [second], "ṭālīṭan", etc. The enumeration can go on as long as there are items in the list.

5. Amplification

Related to the activities discussed in (6) above is that of amplification. This is essentially an enumerative relation but with a different organisational role. The text sequences are listed in a particular order, where the comparable items are listed first and the most prominent one is delayed until the end. The shift from the

comparable to the most prominent item is signalled by connectives, the most frequent of which is "even" in English and "bal" in Arabic.

There are two interlingual variations in the utilisation of this relation. The first is qualitative and refers to the semantic space of "even" as compared to "bal" (see details in 9.2.2.2.3). The Arabic connective is capable of signalling a sharper and more abrupt shift than its English equivalent. The other variation is quantitative. Although connectives of amplification comprise a set of 204 tokens in English, slightly higher than its Arabic counterpart, which consists of 178, it represents a much higher share in additive connectives. The English set stands for 9% of additives while the Arabic set corresponds to 2%.

6. Continuity

A further distinctive interlingual variation concerns the vital role of additive connectives of continuity in sustaining cohesion in Arabic compared to English (cf. the detailed discussion in 6.4.7 and 9.2.2.2.5). There are two basic organisational roles that connectives of continuity are involved in. The first is related to bridging gaps and the second is concerned with connective support. Both have been touched upon in 10.2.3 above when discussing features of text sequencing pertinent to English and Arabic.⁽⁴⁾

10.3.3.4 Extension via Comparison

10.3.3.4.1 Textual Role of Comparison

Connectives of comparison, similar to those of additivity, signal

a type of textual organisation that relies on strategies of extension. It differs, however, from additivity in the type of conceptual motivation that underlies its utilisation. While activities of additivity are involved in cumulating blocks of knowledge and may therefore be termed "extension proper", comparison imposes a type of extension that is based on measurement and discrimination. Basically, a text sequence is extended when it is conceptually delimited by being weighed against or correlated to another (see detailed account in 6.5). This involves one of two functions. Either comparison is made for the purpose of establishing similarity (or otherwise) without the requirement for assessing its intensity or extensity, or it is achieved for specifying comparative size, expanse or proportion. In either case, connectives assist in upholding comparison and indicating its type, scope and direction of emphasis.

10.3.3.4.2 Comparison in English and Arabic

In general, connectivity via comparison plays a small role in text organisation compared to that of additivity. However, as the statistical profile of comparative connectives confirms (see 8.4.1 and 8.4.2; see also 9.2.3.1), this role is stronger and more prominent in English than in Arabic text actualisation. The figures in the profile show that in the English corpus the category of comparative connectives constitutes 4.6% of all connectives while in Arabic they represent 1.6%. Further, the size of comparative activities signalled by connectives represents 19.5% of total activities of extension in English but only 3% in Arabic.

A closer inspection of the text-creating role of comparative connectives in each corpus reveals some interesting interlingual variations. These are summarised below (more details are given in 9.2.3.2).

1. Similarity

a. One main variation concerns the textual behaviour of connectives of similarity in each language. English connectives are of two types. The first introduces textual activities dominated by progression, e.g. "as", "as if" and "as though". The second is involved in activities dominated by regression, e.g. "similarly", "equally" and "likewise". In Arabic, connectives constitute one type: all dominated by progression. This limitation affects text linearisation in two ways, both being characteristic of connectives of similarity involved in progression. The first concerns the imposition of "immediate" range of operationality (see 9.3), while the second is related to the tendency towards tight grouping of antecedent and subsequent propositions (see more details in 9.2.3.2.1).

b. A further variation concerns more extensive typology of subfunctions that English connectives of similarity signal in comparison to Arabic. The subfunctions (see 6.5.2) are better distributed in the corpus and hence manifest a characteristic efficiency of operationality. Arabic connectives of similarity are dominated by two subfunctions that are intensively distributed.

2. Degree

Another main variation concerns connectives of degree. In

English, these connectives are capable of signalling three types of comparisons: equality, inequality and indeterminate degree. In contrast, Arabic connectives are not capable of signalling equality in the same way English connectives do. Moreover, connectives signifying indeterminate degree have a small frequency of occurrence in the Arabic corpus. To express these two types of comparison, Arabic resorts to the use of embedding or complex prepositional phrases.

10.3.4 Variation

10.3.4.1 Operational Scope

This type of connectivity represents a textual strategy whereby a text sequence is introduced in total or partial replacement to the current one. Rhetorically, variation subsumes activities that assist the text producer to interrupt the patterns of textual development in order to present variant arguments or counter-assertions. The interruption may create discontinuity, but its impact is normally reduced via the use of appropriate connectives. These promote cohesion while signalling various activities of variation: modifying, rebutting, subtracting, replacing, opposing, contradicting, contravening, diverging, waiving, contrasting, making concessions or selecting alternative options.

Variation differs from extension in that the latter develops text by ordering, appending, listing, clarifying and/or amplifying. These activities assist in producing a carefully protracted conglomeration of sequences that can further be shaped into a

continuously developing pattern, the elements of which are clearly ordered and related. Variation, on the other hand, suspends this development and initiates an alternative that reflects a variant topical option, a possible replacement, a tentative divergence or direct opposition, or that acknowledges a criticism, a shortcoming or the awareness that something has been omitted or only partially stated.

Connectives signalling variation can operate on sequences of various lengths and complexity. They may connect clauses, sentences, paragraphs or larger "chunks" of text, and can, therefore, have any operational range in the text. This makes them highly effective devices for organising text components.

10.3.4.2 Types of Variation

Variation as a textual strategy that is signalled by connectives is achieved by two types of connectivities: adversativity and alternation. The textual role of each type is discussed in detail in Chapters 6, 8 and 9. We shall therefore restrict the discussion below to outlining the major contribution of each type towards text development.

10.3.4.3 Variation via Adversativity

10.3.4.3.1 Textual Role of Adversativity

Adversative connectivity constitutes a set of powerful text-creating strategies. In general, it assists the text producer to halt what sounds like a "positive" text development and commences a "negative" development whereby he either expresses a direct

opposition or makes an exclusion to the foregoing statements, or discusses the topic in terms of what it does not mean or does not give rise to. The most vital rhetorical aspect of the connectivity is the incompatibility of views, comments, outcomes or results made in the second development (that is, expressed within the subsequent) to those discussed in the first (within the antecedent). This aspect contributes to establishing focus and building up textually effective sequences.

10.3.4.3.2 Adversativity in English and Arabic

1. Connectivity via adversativity plays a major role in text creation in English. This is confirmed by quantitative evidence drawn from the corpora (see details in 8.10 and 9.2.10.1).

a. The set of adversative connectives represents the largest category in the corpus in terms of absolute and relative frequencies. The set comprises 2,404 tokens that represent 25% of total connective tokens. In contrast, the adversative set in Arabic comprises 1,692, representing 10% of the total connective mass. Extrapolation, on the basis of growth, yields larger differences. An English corpus of a million word length is expected to contain 10,082 adversative tokens, while an Arabic corpus of a similar size is estimated to comprise 5,398.

b. Another characteristic of English adversative connectives is their higher rate of repetitiveness in the corpus compared to Arabic.

2. A third characteristic is peculiar to connective range of

operationality across both corpora (see details in 9.3.12). Observation confirms that there is a stronger tendency in English than in Arabic towards utilising adversative connectives for achieving short and long ranges of connectivity. In other words, English adversative connectives are used more often for relating sentences and larger text segments, and so help more effectively in text creation, than their counterparts in Arabic.

3. Related to point 2 is the relatively low frequency of Arabic equivalents to the English connectives "however" and "yet". These two connectives are a) capable of having all types of operational ranges, and b) characterised with relatively high frequencies. In short, they play a powerful role in the application of strategies of variation for organising text.

4. Another equally important difference lies in the strong tendency in Arabic text organisation to use additivity in an adversative environment, thus utilising a combination of extension and variation to establish connectivity. This is confirmed by the following observations (see also 9.2.10.2):

a. Adversative connectives in Arabic require support for relating sequences. This is provided by the connective "wa" which establishes continuity and, at the same time, reduces the full impact of adversativity by conflating it with additivity.

b. The connective "wa" may often be used in an adversative, both antithetic and contrastive, function, and, in this position, conflates an additive with contrastive meaning.

c. A characteristic pattern of some Arabic adversative

connectives, which is not present in English, is the utilisation of an accompanying additive connective to create focus. In this case, the connective, such as "'ammā", is unable to support connectivity by virtue of its own role and requires the connective "fa", positioned after the connective, to introduce the subsequent proposition.

5. Added to these differences are the peculiarities of textual patterning that characterise Arabic connectives in comparison to English. These involve positioning (mobility vs. immobility), polarity (negative vs. positive construction), and type of clause pattern (nominal vs. verbal) (details with exemplification are given in 9.2.10.2).

The textual tendencies discussed above, combined together, point to different application of the strategies of variation in text creation in each language. To put it differently, they represent two variant schemes of text linearisation that exhibit differences in the way text segments are sequenced and salience is created and maintained.

10.3.4.4 Variation via Alternation

10.3.4.4.1 Textual Role of Alternation

The main role of alternation as signalled by connectives is to indicate the presence of possible options in the textual world. These can be facts, events or courses of actions, all possible in the textual world, but only one is valid (see 6.6 for a detailed discussion). That is, the knowledge configurations in the textual

world are arranged in such a way that the options are alternatively the same or similar in regard to their environment and hence one, not all, is accessible. This is an exclusive type of organisation which can be of two types. The first is necessary exclusivity where the options are mutually exclusive due to conceptual/logical inconsistencies or contradictions and therefore connectives signalling the relation assist in establishing a correct textual structure. The second is accidental, where the alternative configurations are not in themselves logically contradictory, but relative to the context only one is valid (see 6.6.2 for more details and examples).

10.3.4.4.2 Alternation in English and Arabic

1. One characteristic difference between English and Arabic alternatives is the relatively higher repetitiveness of alternative connectives in English. For despite the smaller number of tokens in English, their relative frequency is higher, indicating a comparatively more intensive utilisation of this relation for structuring text sequences.

2. Another main characteristic difference of this type of variation concerns the semantic scope of the alternative connectives in each language. Close inspection suggests that English alternatives ("or", "either ... or") are fundamentally "exclusive": either one of the alternative configurations is true, but not both. In general exclusivity is a powerful means of creating variation in the conceptual/rhetorical organisation of text and, in English, it exhibits a sharper definition of relation. In Arabic, although the

alternative connectives are frequently exclusive and therefore their semantic scope coincides with that of English alternatives, they may also be utilised to signal a more "inclusive" meaning, combining conceptually similar propositions that are both accessible in the textual world, a meaning that is essentially additive. In other words, Arabic alternatives, as the discussion in 9.2.4.2 establishes (see also Figure 9.28), manifest an inherent capability for operating on a wide semantic continuum where, on one end, the alternative relation is exclusive, and on the other it is inclusive, largely more additive than alternative. At this point, the connectives signal a textual fusion of two organisational strategies: extension and variation. The continuum can be expanded further and the relation becomes of the additive/coupling type (see 6.4.8).

3. Alternative connectives in both languages are characterised with an immediate operational range, combining mainly clauses to make up larger clause-complexes (see Table 9.11 and the discussion in 9.3.6). They, therefore, exhibit a relatively limited organisational role in relating larger text sequences within each corpus.

10.3.5 Elaboration

10.3.5.1 Operational Scope

Elaboration subsumes strategies whereby one text sequence, i.e. the subsequent, develops the content of another, i.e. the antecedent, by further specifying it, identifying its details,

evaluating its components and/or assigning points of departure. The operations aim ultimately at creating text through providing a further, more sophisticated characterisation of a particular conceptual configuration. The role of connectives is to mark and itemise the type of elaboration intended and initiate the various relevant operations.

One basic motivation for utilising activities of elaboration is to establish salience. That is, all activities and decisions regarding the subsequent text sequence are intended to amplify perceptual space (the totality of what is perceived at any moment) in such a way that the conceptual content of the antecedent impinges more strongly on the senses. This is accomplished when the content of subsequent text looks back, via activities of regression, to the content of the antecedent and reflects, clarifies or refines its conceptual scope.

10.3.5.2 Types of Elaboration

Connectives are capable of signalling various types of elaborative activities. These can be grouped into two major types: reformulation and orientation (refer to the relevant discussion in Chapters 6, 8 and 9). Each type represents a different strategy that involves a different set of textual operations, resulting in a specific type of text organisation.

10.3.5.3 Elaboration via Reformulation

10.3.5.3.1 Textual Role of Reformulation

Connectives of reformulation signal activities whereby the

subsequent sequence re-presents the content of the antecedent either by restating, explaining, exemplifying or summarising it. The aim is to proffer a clear characterisation of the informational content of the antecedent and by doing so helps develop it within the text (see 6.7.1).

There are three main types of reformulatory categories:

1. Restatement: This refers to re-introducing the content of the antecedent (see details in 6.7.2) either by direct paraphrasing, where the propositional content is retained, or by creating a more specific version of it.

2. Exemplification: Here the connective indicates that the subsequent represents one particularised aspect, type or instance of the propositional content of the antecedent (see details in 6.7.3).

3. Summary: The subsequent offers a condensed restatement of the content of the antecedent, usually placed at the end of a phase in the organisational structure of the text (see details in 6.7.4).

10.3.5.3.2 Reformulation in English and Arabic

1. Text organisation in Arabic utilises connectivity via reformulation more intensively than in English. This is a reflection of a pressure in Arabic to repeat, redefine, and rephrase statements in order to create forceful assertion. Combined with other types of repetition, such as parallelism, and additive/coupling connectivity, reformulation in Arabic keeps the text receiver at a definite viewpoint while at the same time

exploring, redirecting or attracting new propositional material to it (cf. Al-Jubouri 1984).

2. Statistically, Arabic reformulatory connectives constitute a larger category, compared to their English equivalents both in terms of tokens and types. Moreover, they manifest comparatively higher rates of repetitiveness throughout the corpus (see details in 8.5.1.1, 8.5.2.1 and 9.2.5.1).

3. Examination of the distribution of each reformulatory category (see 8.5.1.2, 8.2.2.2) shows that connectives of restatement and exemplification constitute larger sets in Arabic with higher repetitive rates than in English. Connectives of summary constitute a small set in each corpus, though their frequency in English is higher than in Arabic.

4. Connectives of exemplification are associated with two variant patterns (see discussion in 9.2.5.2.2). In pattern A the exemplificatory sequence (the subsequent) is a self-contained block of knowledge. Once the relation is established, the sequence content is textually exhausted and development of the antecedent is resumed. In other words, the content of the exemplificatory sequence does not control the informational/rhetorical development of the next sequences in the text. This pattern is favourable in text organisation in English, a finding that is confirmed by its high frequency in the corpus.

In pattern B, the exemplificatory statement is not self-sufficient. It is capable of creating sufficient salience to cause a shift in focus to take place, which then requires more development

in the next sequences. In other words, the content of the exemplificatory statement, which elaborates the main statement of the antecedent, becomes itself a topical one and thus exercises a textual influence on later conceptual/rhetorical development. This pattern is more frequent in Arabic, as observation of the corpus confirms.

5. There is a stronger tendency in Arabic to utilise reformulatory connectives for short range connectivity. In contrast, English connectives show a higher relative (though not absolute) frequency of occurrences as long range connectives. However, this high frequency is balanced by the higher number of paragraphs in English (see details in 9.37 and Table 9.12).

10.3.5.4. Elaboration via Orientation

10.3.5.4.1 Textual Role of Orientation

Orientation is the more dynamic type of the two types of elaboration strategies. Reformulation is, generally speaking, restricted to restating, clarifying and refining the content of the antecedent without a pressure for introducing new conceptual chunks. Hence text development is limited to those activities. Orientation, in contrast, enables the text producer to reconsider the propositional content of the antecedent and introduce a clarification, an evaluation, a generalisation or a confirmation. This type of strategy is utilised whenever the text producer resorts to reasoning or demonstration and therefore requires to imply his attitudinal presence (see details in 6.8.1).

Connectives signalling the relation suggest a phase in the growth of text where a sequence is reinstated, modified, enlarged or confirmed. More specifically, they have two textual roles:

1. Adjustment: This is where the text producer controls the text statements and provides them with the necessary transitions. "Adjustment" subsumes such textual activities as assigning authority, assessment, generalisation, particularisation, introducing points of departure and specifying textual perspectives (refer to 6.8.2).

2. Confirmation: Connectives signalling confirmation enables the text producer to corroborate, ratify or acknowledge the current statements (see details in 6.8.3).

10.3.5.4.2 Orientation in English and Arabic

1. English orientative connectives constitute a larger category both in terms of tokens and types (see details in 8.6.1.1 and 8.6.2.1) compared to their Arabic equivalents. The difference is magnified further when the relative frequencies of the category in both corpora are compared (10.5% in English compared to 4% in Arabic).

2. In general, the category is characterised with a massive type repertoire. However, the size of the repertoire is significantly larger in English compared to Arabic, with higher variegation and lower concentration. Further, English orientatives manifest distinctly higher rates of repetitiveness than their Arabic counterparts (see comparison in 9.2.6.1).

3. The general quantitative profiles of this category confirm that there is a stronger tendency in English relative to Arabic to organise text by utilising elaborative strategies of orientation. The pressure for explicit use of signals and the wide variety of orientative transitions impose a higher variegation of signals in English (refer to the detailed discussion in 9.2.6.2).

In Arabic, pressure for signalling orientative transitions exists, but it is less forceful and less urgent than in English. Most transitions, unless they are intensely sharp, are either left implicit or, subdued via additivity.

4. English orientative connectives manifest a higher extent of mobility in positioning. This creates various types of linear formatting, and choice is determined by the extent of intended salience and the location of the elements to be placed in focus. In contrast, Arabic connectives of adjustment show relatively little variability of positioning. Their normal position is the initial slot in the subsequent sequence.

5. In Arabic, the use of orientative connectives in initial position in the sequence, even when preceded by an additive "wa", creates a textual gap that dissociates it from the rest of the sequence. This is especially true when the subsequent is not a verbal clause (that is, having a verb in a thematic position in the clause followed by the subject). The gap is sufficient to cause disconnection and temporary disintegration. To rectify this problem, text cohesion requires the use of "fa" in an additive role

(see 6.4.6). This characteristic feature of Arabic connectivity creates a fusion of two types of organisational strategies: extension and elaboration.

10.3.6 Enhancement

10.3.6.1 Operational Scope

Enhancement subsumes connectivities whereby the content of a subsequent sequence creates text by enriching or qualifying the content of the antecedent. The operations associated with enhancement are discursive in the sense that they aim at exploring the content of the current sequence, augmenting it and, often, moving out from its topical content.

A basic difference between enhancement and other means of development lies in the way the subsequent is related to the current text sequence. Normally the subsequent sequence considers all or part of the conceptual content of the current text sequence and works out a type of qualification, such as indicating causes or effects, giving explanation or expressing motivations. Alternatively, qualification may involve temporal or spatial associations, which then may expand the propositional content in accordance with the planned goals of the text.

10.3.6.2 Types of Enhancement

Enhancement is a common strategy in text development. It relies on three types of main connectivities: temporality, causality and spatial relation (see details in Chapters 6, 8 and 9). Each has its own distinctive textual role, though they provide similar

contribution to the creation and organisation of text.

10.3.6.3 Enhancement via Temporality

10.3.6.3.1 Textual Role of Temporality

Connectives signalling temporality provides time references that assist in organising various knowledge configurations in the textual world (see 6.9.1). Time references vary according to the semantic scope of each connective type and that is reflected on how propositions are related. Inspection of the behaviour of temporal connectives in the corpus identifies six types:

1. Temporal sequence (or succession): Connectives signalling this relation establishes a temporal order through which text segments are organised (see details in 6.9.2).

2. Simultaneity: Connectives of simultaneity indicate that an event, state or course of action expressed in the subsequent runs parallel to that in the antecedent (refer to 6.9.3).

3. Span: Connectives of this category introduce a proposition that determines the temporal span, whether forward or backward in time, during which another proposition is valid (see 6.9.4).

4. Positioning: Here connectives organise text segments by specifying a point on the time dimension of one and positioning another onto it (see 6.9.5).

5. Temporal Frequency: This is a fusion of a temporal and conditional meaning. Connectives of this category signal the

dependence of the validity of the subsequent upon that of the antecedent (see 6.9.7).

6. Circumstance: This is a variant pattern of temporal simultaneity that has structural and functional properties that make it one of the peculiarities of Arabic (see 6.9.6).

10.3.6.3.2 Temporality in English and Arabic

1. Temporals in General

Temporal connectives constitute a large category in both languages. In English, temporals stand for 13% of total connectives; in Arabic, they represent 9%. In terms of absolute frequency, Arabic temporals make up a bigger set than their English counterparts: 1,494 vs. 1,284 (see more details in 8.7.1.1 and 8.7.2.1). These numbers reflect the significance of using connectives to organise text sequences on a temporal basis. To understand the textual role of temporals in each language, we have to consider the various subcategories, each separately.

2. Temporal Sequence

a. Connectives signalling temporal sequence or succession play an important role in organising sequences in text in Arabic in comparison to English. This is evident in their frequencies in the corpora: the category constitutes 52% of temporals in Arabic and 28% in English (see 8.7.1.2 and 8.7.2.2). This is a reflection of a stronger tendency in Arabic to connect text components temporally whenever their content exhibits an ordering in time, even if it is vague or remote. This may occasionally create fuzziness in the

semantic scope of the relation due either to reduction of functional distinctness or complexity of signalling (on this see 9.2.7.2.1 points 1 and 2).

b. One important organisational feature of temporal sequence is gradation of temporal succession. In Arabic, the connective "fa" points to immediate succession of events or courses of actions, while the connective "tamma" indicates a longer time lapse.

c. The connective "fa" may also be used to organise succession of roles of speakers in a reported conversation in Arabic (see details in 9.2.7.2.1 point 3). A similar explicit signal has no important role in English.

d. Another organisational tendency that is far more distinctive in Arabic compared to English is the multiple use of temporal connectives, particularly "fa" or "tamma", to establish a chain of successive events or courses of action (see 9.2.7.2.1 points 3 and 4).

3. Temporal Simultaneity

a. There is a stronger tendency in English to signal temporal simultaneity of events or courses of action by the use of connectives. This is evident in the higher frequency of connectives of temporal simultaneity in English (15% of temporals) compared to Arabic (4%). In contrast, Arabic displays a lesser textual pressure for using connectives to organise text components on the basis of temporal simultaneity.

b. There are two types of simultaneity, one designated simultaneity "proper" and the other "abutting" simultaneity (see details with examples in 6.9.3). While English utilises connectives to signal both types, depending on rhetorical requirements, Arabic favours the "abutting simultaneity" type and only a small number of connectives are used to signify relations of simultaneity "proper". The latter type of relations is more frequently expressed via circumstantial constructions and signalled by "wa".

c. English connectives signalling simultaneity proper enter into a number of organisational structures. Three have been identified: span-span, event-span and event-event simultaneity. In Arabic, there is a tendency to signal event-span simultaneity and its related variants (see 9.2.7.2.2).

d. A further organisational difference is created by the structural nature of the connective itself. In English, the most frequent connectives are subordinators: "as" and "while". In Arabic, the most frequent connectives are phrasal with anaphoric reference. Each type influences text linearisation differently. This is reflected in the way text segments are grouped and sequenced and the manner in which salience is created (see 9.2.7.2.2 point 2).

4. Temporal Positioning

a. There is a strong tendency in English to utilise connectives to signal temporal positioning. This is confirmed by the high frequency of this category in the English corpus (49% of temporals) compared to Arabic (27%).

b. The main organisational difference is related to the position of Arabic temporal subordinators. When in initial position in the sentence they create discontinuity, which impairs its cohesiveness with the previous stretch of text. To rectify this problem, an additive connective, usually "wa", is used. A similar textual feature does not occur in English (for more details on this and also other variations see 9.2.7.2.3).

5. Temporal Span

a. Inspection of connectives signalling this relation in both corpora confirms that while English connectives may signal "backward" (past) or "forward" (future) spans with approximately equal probability (mainly through the use of the subordinators "since" and "until"), Arabic connectives display a very high probability for expressing "forward" span (mainly through the use of "ḥattā" and "'ilā 'an").

b. Another difference arises from the subtleties of signalling associated with "until" as opposed to "'ilā 'an" and "ḥattā" (see details in 9.2.7.2.4). The connective "ḥattā", for instance, can cover the full functional scope of "'ilā 'an" and "until" and may be extended until the meaning becomes causal (see Figure 9.47).

c. Like connectives of temporal positioning, Arabic connectives of span creates discontinuity when they occur in initial position and are therefore supported with an additive connective.

6. The role of circumstantial "wa" is a characteristic feature of

Arabic (see a detailed account in 6.9.6 and 9.2.7.2.5). The main organisational role is the association of a particular situation, expressed in one text segment, to an event, state or course of action expressed in another segment. This is a common type of connectivity and the frequency of the connective represents 8% of temporals in Arabic.

7. Comparison of range of operationality of temporal connectives in both corpora reveals a tendency that is prevalent in both corpora for immediate range connectivity. However, Arabic temporals also show a tendency towards short range connectivity while their English equivalents manifest stronger tendencies towards medium and long operational ranges.

10.3.6.4 Enhancement via Spatial Connectives

10.3.6.4.1 Textual Role of Spatial Connectives

The basic textual role of this category of connectives is the provision of spatial orientation whereby one text component (the subsequent) is related to another (the current sequence or the antecedent). Normally the relation is that of spatial position, i.e. the content of one proposition is related to (e.g. occurs in) a point included in the conceptual content of another. The connectives, therefore, lack the varied functionality and techniques of organisation that temporal connectives manifest.

A variant role is where the spatial relation is figurative. That is, the connective relates one proposition to a figurative location constructed within another.

10.3.6.4.2 Spatial Connectives in English and Arabic

1. Spatial connectives constitute a small category in both corpora. In English, the category comprises 50 tokens representing 0.5% of token mass in the corpus, while in Arabic it consists of 40 tokens, corresponding to 0.2%. The difference in size is indicative of wider utilisation of this type of connectivity in English relative to Arabic for organising text components and sustaining their cohesion.

2. Most occurrences of spatial connectives in both corpora display an "immediate" operational range. However, English connectives are utilised more frequently than their Arabic equivalents to achieve short, medium and long ranges of connectivity.

10.3.6.5 Enhancement via Causality

10.3.6.5.1 Textual Role of Causality

Causal connectivity is one of the most effective strategies for text developments. Connectives can effectively be used to set up a nexus whereby the content of one sequence is explored, amplified or clarified. Development can be limited in the sense that a cause, effect, inference, or condition of the propositional context of a sequence is stated for a local clarification of the content of the antecedent and is therefore not developed further. Alternatively, it may be extensive when the expression of causality introduces expansion to the topical content of the antecedent and moves out from it. The choice is governed by the plan and goals of the text

(on the nature of the relation refer to the comments in 6.11.1).

10.3.6.5.2 Causality in English and Arabic

1. Text organisation in Arabic makes an extensive use of causal connectives for text development. This is confirmed by the higher frequency of this category in the corpus. Arabic causal connectives constitute the second largest category in the corpus, comprising 3,778 tokens that represent 22% of total connectives in the corpus (see more details in 8.9.2.1). Likewise, English causal connectives play a significant role in text organisation though to a relatively lesser degree than additives or adversatives. The category comprises 1,888 tokens representing 20% of total connectives in the corpus (see 8.9.1.1). A better understanding of the role of causals can be attained by examining each of their five subcategories.

2. Cause/Reason

a. Connectives signalling cause/reason constitute 33% of causal tokens in Arabic and 26% in English. In terms of absolute frequency, this category of causals comprises 1,259 connectives in Arabic and 484 in English (see also 8.9.1.2 and 8.9.2.2).

b. A main difference in textual role of this category of causals across the two languages lies in the organisational force of "fa". This connective assists in establishing one of the frequent patterns of paragraph development in the language. The basic pattern starts with a topical statement of whatever structural complexity. Then "fa" is used to introduce a statement (or set of statements) that explores the topical one and moves out from it (see 9.2.9.2.1 point

2 for exemplification). This pattern may be reiterated in the paragraph in a manner not admissible in English text organisation (see 9.2.9.2.1 point 3 for details and examples).

3. Result/Influence

a. Connectives signalling this function play a more significant organisational role in Arabic in comparison to English. This is confirmed by the large size of the category in Arabic, which comprises 1,460 tokens representing 39% of causals (see also 8.9.2.2). In contrast, the English set consists of 581 tokens representing 27% of causals (see 8.9.1.2).

b. One characteristic differential feature is the tendency of English connectives to signal a clearly planned causal relation. In contrast, Arabic connectives are more likely to signal indeterminate and remotely related causality (see details in 9.2.9.2.2 point 1; see also Figure 9.58).

c. A common pattern that involves the use of a causal "fa" involves the expression of an assertion as a viewpoint followed by one or more related questions that explores its validity or consequences. The question (or set of questions) is usually introduced by "fa" which assists in maintaining topical unity within the pattern (see 9.2.9.2.2 point 2).

4. Condition

a. Connectives signalling condition plays a far more extensive role in text organisation in English compared to Arabic. This is

evident in their high frequency in the corpus. The category represents 38.5% of causals in English in comparison to 15% in Arabic.

b. Arabic conditional connectives operate under certain constraints that affect their organisational role. First, a conditional connective is more likely to occur medially than initially in the sequence. That is, it is involved in a pattern where statement B (the consequence) precedes statement A (the condition). If a connective occurs initially in the sequence (in which case the conditional statement precedes the consequence), the connective creates discontinuity with the previous text component. This is rectified in two ways: externally and internally. Externally, the sequence is connected to the previous text component by a supportive connective, such as "wa" or "fa". Internally, the conditional statement (statement B) is combined with statement A by "fa" or less frequently, "la". The choice is determined by a number of structural factors including type of connective (see details in 9.2.9.2.3 point 1).

c. Arabic has a larger type repertoire that are utilised to discriminate specific shades of condition. English utilises "if" as a general-purpose conditional connective (see 9.2.9.2.3 point 2).

5. Degree-magnitude

a. This is a relatively small category in both corpora, though it is larger in English (representing 4% of causals).

b. The connectives "so ... that" and "such ... that" organise

text sequences in a manner that is characteristic to English. The two connectives direct focus to a particular concept within statement A and relate the consequence to it (see discussion in 9.2.9.2.4). Arabic connectives, in contrast, look back at the entire conceptual content of statement A intensifying it to an extent where the consequence becomes valid.

6. Purpose

This is a large category in Arabic, representing 12% of causals. In contrast, English connectives of purpose correspond to 1.3% of causals. This indicates that English uses other means of signalling purpose than connectives (one of which is the use of the infinitive).

10.3.7 Concluding Remarks

The discussion on the previous sections has concentrated on the various types of connectivities and the operational scope of each in both languages. A number of differences have been observed and identified in the way sequences are related and combined into larger textual components. These differences, collectively, characterise cohesion and text development in each language.

For instance, we have argued that the most prevalent method of text organisation in Arabic relies on the utilisation of strategies of extension to create text. One sequence is given first and then expanded by adding another to it. In English, text organisation is influenced by the use of strategies of variation and then extension.

Another difference is related to the way strategies of

elaboration are utilised. Arabic text is under pressure to review, define and restate previous sequences. English text is under pressure to provide points of orientation whereby the subsequent sequence adjusts the perspective of the current or previous sequence.

A third difference lies in the way strategies of enhancement are utilised. In expressing temporality, English has a tendency to anchor at temporal points within the previous text sequence and starts to develop the text from there. Arabic has a tendency to develop text on the basis of temporal succession. Other strategies that are operational to a wider extent in English than Arabic is the use of temporal simultaneity and backward span to organise sequences temporally. Arabic, on the other hand, exhibits a stronger tendency to express attendance of circumstances and situations.

Strategies of enhancement also differ in the way causality organises text sequences. Arabic tends to exploit expression of cause/reason or results/inference far more readily for topical development within or across the paragraphs. English has a stronger tendency for exploiting conditional statements in organising text sequences.

There are other more intricate and more sophisticated variations in functionality that concern positioning and its effect on the linear grouping and sequencing of text components, on achieving salience, on the extent of specificity (as opposed to generality or indeterminateness) of signalling, and on the way text integration and continuity is maintained.

We would next like to use these major variations to offer some pedagogical suggestions for preparing EFL material for teaching written composition.

10.4 Implications for Teaching EFL Writing Skills at the Undergraduate Level

10.4.1 Preliminaries

The focus of attention within this brief section is on the provision of some pedagogical guidelines that help translate the general findings of this study into classroom procedures. Detailed analysis and synthesis of practical exercise design lie beyond the scope of this Chapter and are intended to be taken up as one of several follow-up projects (see 10.5). Our immediate concern is to make a number of suggestions that can be of two related uses: a) to direct the design of material to supplement existing writing manuals through providing a missing student-oriented component of the teaching operation, and b) to assist the design of better-equipped writing programmes for the Arab learner in general and our target learner, the university undergraduate (see Ch. 1) in particular.

For convenience, we divide this part of the study into four brief sections. The first section (in 10.4.2) discusses some aspects of an effective writing program, some general in nature while the others more specific. The second section (10.4.3.1) summarises a few requirements of material design that make a background for later discussion. The third section (10.4.3.2) is central to the discussion and includes discrete instructions on design. The last section lists a few types of practice exercises.

10.4.2 The Writing Programme: an Introduction

One of the major aims of the undergraduate EFL writing programme is to help the student to have complete control of the linguistic mechanism that enables him to make competent expression of facts, views, events and courses of action. The most central instructional activity by means of which the structural and textual forms and patterns are organised into a cohesive/coherent whole in relation to a particular topic is that of composition. We would like to note that the linguistic mechanism by itself, albeit necessary, is not sufficient for the production of a composition that satisfies both the constitutive and regulative principles of textuality (see Ch. 3). Other factors, such as topical awareness and motivation, play major roles in shaping the content and organisation of the text and have therefore to be catered for in the design and implementation of the instructional programme. These factors, however, do not constitute the focal concern of this section and therefore a detailed consideration of their effect will be bypassed.

We propose that there are two (probably among several) fundamental aspects of teaching written text production. The first is lexico-grammatical and covers the efficient, effective and appropriate choice of lexical and grammatical elements to be utilised in the expression of the text content. The second concerns the linearisation of textual components (see Ch. 3 for a detailed discussion of linearisation) and covers textual choices (for instance topical/rhetorical selections) as well as the process by which they are materialised into a finished product.

The two aspects are inseparable and this classification should be taken as a convenient procedure for inspection and assessment. In text creation, there is a close interaction between lexico-grammatical choices and those that determine text organisation. This reflects itself in the various activities involved in linearisation: grouping text segments, sequencing them and achieving various degrees of salience.⁽⁵⁾

The learner then has a dual task to achieve. He is required, first, to write linguistically accurate sequences in English (in so far as it is inaccurate it ceases to be English) and, second, to put the sequences together in a manner that satisfies English textuality. This task demands that the learner should develop a linguistic repertoire which he relies on in achieving the first aspect of the task and to be able to exploit its elements to achieve the second.

But despite the close interaction between the two aspects of text production, classroom instruction has been primarily concerned with developing the first, i.e. lexico-grammatical, aspect.⁽⁶⁾ This is reflected in the intensive and extensive practice given in the various instructional manuals (see critique in Ch.1), and the fact that the learner, in general, is left with lesser guidance on how to put the various sequences together in a coherently organised text. Since various aspects of organisation are language specific, which this study has confirmed, it follows that greater attention, preparation and assistance is required to enable the students to write competently.

10.4.3 Some Guidelines for Designing a Writing Programme

10.4.3.1 Requirements

1. Itemisation and Integration

In designing a writing programme, two seemingly conflicting features are required: a clear itemisation and close integration of the various types of instructional activities. Itemisation is essential in the sense that activities have to be structured in a way that is responsive to the goals of the writing programme.⁽⁷⁾ These activities reflect a general gradation of the syllabus and may vary in complexity and typology of classroom procedures. Further, activities may also vary - through constant modification, reformulation and improvement - to match the feedback that results from assessing the learning/teaching experience.

The second feature of a writing programme concerns the requirement of integration. Instructional activities should manifest two types of integration: internal and external. Internal integration refers to the progression within the activities and the reinforcement that one gives the other. External integration has two aspects. The first refers to the relationship between the writing activity and other types of EFL activities, such as reading comprehension, classroom (group) discussion or laboratory or computer-assisted practice. The second refers to the relationship between the activity and learner's experience and motivation.

2. Flexibility

The instructional activities in an EFL writing programme aim at

achieving a series of gradual approximations towards the norms of English textuality. Although there exists a strong element of progression that runs through the various activities, so that writing gradually becomes less controlled and more expressive, the programme should be flexible enough to allow a "regressive" manipulation of controlled writing strategies for helping advanced learners improve their expressive skills. That is, even at the later stages of the course, an advanced learner will still need some controlled writing tasks for reinforcement or remedial work, and the course has to accommodate this type of tasks.

10.4.3.2 Some Pedagogical Suggestions

As argued earlier (cf. Chapter 1 and this Chapter), one of the primary aims of an advanced EFL writing programme for the Arab undergraduate is the development of the capability of producing a text that satisfies principles of textuality that are pertinent to English. This task has several implications of which two are evident:

- a. The need to acquire competence in producing an English type of textual organisation.
- b. The need to develop proficiency in utilising connectives to establish various types of connectivities.

Clearly, it is not easy to teach all the functional and organisational conventions of English written text within one course, and hence careful selection is necessary. However, one of the most problematic areas associated with designing a course of

this nature is the order of priorities on which such pedagogical operations as selection, grading and presentation of instructional activities can be based. This problem situation is aggravated by the fact that applied linguistic studies available, at least at present, are unable to provide a satisfactory answer. Available publications on EFL teaching of written skills are largely do-it-yourself attempts. They are based on individual teaching experience which, though helpful in some respects, are not fully relevant; some even fail to reveal an EFL teaching concept from which general didactic and methodological conclusions could be derived (see further discussion in Chapter 1).

However, in the light of the evidence we have gathered we are in a position to suggest a number of practice types that assist, when incorporated in the writing programme, in directing the teaching operation efficiently and systematically towards the achievement of its desired goals. Two points, however, have to be left to the decision of the material designer or, more appropriately, to the classroom practitioner. The first concerns the size of the practice that is to be given, i.e. the amount of work devoted to that type of practice, including classroom activities and any supplementary work aimed at reinforcing it. The second point is related to the complexity of the exercises that are to be designed and administered. Obviously, learner-centred instruction requires for an optimal effect an individualised type of practice that can efficiently meet the demands of the teaching situation.

A. Organising Text

1. Practice is needed for teaching the learners to organise text in components, or "chunks", of smaller size than those familiar to the students in their native language. This involves the following tasks:

a. An underlying segmentation of conceptual chunks is a necessary pre-requisite for an efficient surface text segmentation. The learner has to be aware that an alternative procedure for developing a text world is by itemising it into conceptual configurations that reflect i) more discernible component parts and therefore a lesser degree of merge and complexity, and ii) clearer presence of defined limits and dividedness.

b. On the surface level, task (a) can be approached by teaching the learners to map conceptual configurations into textual units (sentences and paragraphs) that correlate with the demands of a specific communicative situation and manifest the following characteristics:

i. Boundedness: Textual units such as the sentence or paragraph should have a clearly defined rhetorical function that contribute to the thematic, temporal or spatial orientations of the text. Once the function is satisfied, the unit, as a rhetorical construct, reaches its threshold of termination and another is initiated for a related, though essentially different, function.

ii. Complexity: The rhetorical function should determine the extent of length and complexity of the textual unit. But, in

general, the learner should be taught to avoid making frequent use of excessively complex sentences or producing extensively lengthened paragraphs. Such units, though a regular feature of Arabic text organisation, are not equally frequent in English.

iii. Dividedness: Since the rhetorical function of the textual unit delineates its boundaries, the unit displays an extent of discreteness that is orthographically marked. Hence the efficient use of initiators (such as capitalisation and indentation) or terminators (such as the full-stop). Teaching clear indication of dividedness is a vital step towards helping the learners to approximate English text organisation.

2. Practice is needed for teaching the appropriate conventions of textual sequencing. This in particular refers to the appropriate use of means of structuring text units and relating them in a cohesive whole. The most prominent operations include use of various cohesive strategies to perform textual juxtaposition, regression and progression. A number of tasks are particularly relevant:

a. Learners have to be taught to reduce the high extent of explicitness of signalling connectedness via the use of connectives. While Arabic is heavily dependent on explicit signalling of both cohesive relations and sequencing operations, and therefore connectives play an important role in organising text, English exhibits less tolerance with excessive explicitness of signalling. The learners have to be conscious of the fact that sequencing operations can be performed by various means (other than

connectives) and that relations between textual components may be left implicit. On the other hand, the learners should be trained on how to provide sufficient cotextual and contextual information that helps readers infer these relations.

b. Practice is needed for teaching the appropriate and diverse signalling of textual relations. Connectives serve to indicate the structure of thought, and their diversity and extensivity of use can define effectively and efficiently the movement of the thought expressed: for instance, introducing an illustration, making a concession, expressing a contradiction, or announcing a consequence or an inference. While Arabic connectivity resorts to a limited set of connectives that are used intensively, English connectivity makes use of a wider range of connectives that are utilised under more restricted conditions of choice for establishing text sequencing. Hence variability of connectives and the appropriate use of their patterns constitute a major teaching task in a writing programme. In the absence of a formula for using connectives in a rhetorically effective context, the exposition of connectives in this study can give both the teacher and textbook writer ample opportunity for teaching textual patterns of use. Of course, pedagogically speaking, the point is not so much to encourage learners to repeat accepted devices, though this is of great assistance in EFL teaching, as to help them master the principle - the necessity for maintaining text sequencing via textual connectivity while indicating various shifts in meaning and emphasis.

c. Related to (b) is a type of practice that gives the learner the awareness that English connectives, unlike their Arabic

counterpart, often do not require support with other connectives to establish textual continuity. A connective in English can uphold continuity by virtue of its semantic/rhetorical functioning; in Arabic, most connectives, particularly adverbials, while capable of signalling textual relations, fail to maintain text continuity, thus causing considerable disturbance to stability, unless they are supported themselves with another connective, usually an additive.

B. Establishing Connectivities

1. An essential component of a writing programme should be devoted to training the learners to avoid excessive use of extension as a strategy in text development. This task can be broken into a number of specific but closely related ones:

a. Practice is required to train the learner to reduce their utilisation of additive connectives to establish textual link. This means that learners should, in the first place, be enabled to group and organise conceptual configurations so that reliance on additivity is reduced in favour of other types of linkage. This tremendous endeavour should start at an early stage in the programme, aiming, through gradual training, to acquire sensitivity in developing a text world model that is at variance with the one that the learner is used to.

b. The text world model that is to be developed should give a substantial reliance on variation as a strategy of text organisation. Variation demands a structuring whereby a particular conceptual configuration is set in close proximity with other

incompatible ones. On the surface level, either a statement is first made and then a concession is admitted, or an argument is made and then a counter-argument is announced. This involves teaching the learner to view a statement from two opposing perspectives, and use the appropriate connective to establish the link.

c. Learners should also be trained to reduce excessive use of connectives indicating paraphrase or reformulation. These two types of linkage are common in Arabic, reflecting a textual pressure to look back and define or restate a concept or a proposition. Learners have to be aware that such a pressure exists to a far lesser extent than their experience permits.

d. Practice should be provided to enable the learners to reduce reliance on parallelism for text-creation. Parallelism is a common feature of Arabic text organisation, but its use is relatively restricted in English. Hence excessive utilisation of connectives, particularly additives, for producing parallelistic forms can give an English text an unfamiliar organisation.

e. Related to (c) and (d) is a type of practice that aims at avoiding excessive use of isomorphic propositions. This is the additive relation of coupling, which is frequent and favourable in Arabic at various textual ranks (words, clauses, etc.). English textuality, generally, makes a sparse use of isomorphism, unless, of course, the choice is rhetorically dictated.

f. Careful, graded practice is also required to train learners in the appropriate utilisation of the strategies of enhancement pertinent to English text organisation. This involves helping the

learners to render i) a stricter expression of causality via connectives than is permitted in Arabic, ii) a better distinction of the textual patterns of connectives of condition and purpose, iii) a lesser degree of reliance on connectives of temporal succession for developing text components and iv) a more subtle distinction in the patterns of temporal connectives of span and simultaneity.

g. Practice is also required to support the learning of more varied patterns for the expression of enumeration. Learners should be aware of the method of utilising connectives for listing the various conceptual components available in the text world and organise the surface text accordingly.

h. Intensive practice is needed for teaching the learners the appropriate patterns of orientative connectives. These connectives signpost the various text components, thus providing accurate directions and putting each textual component in its appropriate perspective. Since the typology of textual directions is diverse, it follows that these signposts are extensively varied. Hence the learner should a) have at his disposal a variety of connectives to use whenever there is a requirement to signpost, b) be able to generate signposts whenever provision of a perspective is deemed textually convenient - i.e. assists in achieving higher efficiency or wider effectiveness.

10.4.4 Typology of Exercises

The types of practice that have been suggested are capable of improving through carefully graded exercises, the textual quality of

learners' written composition. Types of exercises can vary, and their use is governed by the specific demands of the teaching situations. In general the textual use of connectives can be practised in three major types of exercises:

1. Preparation exercises

These exercises can be designed independently of other instructional activities and are used to focus specific attention on a certain task, i.e. the use of a particular problematic pattern. They aim at preparing the learner for other types of reading/writing activities and can, therefore, be regarded as an integral part of a wider teaching task.

Preparation exercises can be classified into a number of types that are now well known to practitioners. Some effective types are:

- a. Sentence combining: The use of this type of exercises have proved to be significantly serviceable for improving writing skills of native speakers of English (cf. papers in Daiker et al. 1979 and McQuade 1986). It has also been used with non-native speakers to improve writing proficiency (Angelis 1975). When well designed, the exercises can give the learner the necessary training in performing appropriate textual grouping and sequencing at various levels of complexity. Most importantly, they can give the students training in making choices that are contextually compatible and textually efficient.

- b. Completion: This is a useful type of exercise that is gradable according to levels of complexity. Completion can either be

controlled by the provision of a list of choices, or expressive, i.e. left to the learner to express the most appropriate choice.

c. Transformation: Various versions of this type can be used for teaching connectivity. One effective type is to ask the learners to transform or modify a text sequence that is integrated in accordance with one strategy into another using a different (but specified by the exercise) strategy.

d. Expansion: This, again, can have various versions, depending on the complexity of the task and sophistication of the learner. The exercise may require expansion of a particular statement by integrating various components to it via the use of connectives. Connectives can either be provided in a list and the learner can make a choice, or provision of the appropriate connective can be left to the learner to select and incorporate (probably via peer work and under the teacher's supervision). An alternative, more advanced version, is to supply a list of vocabulary items that the learner can incorporate into making a complete textual sequence (a sentence or, better still, a paragraph).

2. Exploitation exercises

These are used in conjunction with other activities: a group discussion, a listening comprehension exercise, a computer exercise or game, a video film or programme, or a reading comprehension passage. These activities are useful for providing the necessary content that gives the learner a conceptual framework to operate within. Exploitation exercises normally extend and reinforce the activities they are associated with and reflect a close

internal/external integration. The same types of exercises that have been suggested under "Preparation Exercises" can be applied here. Their degree of sophistication depends on such factors as the learner's level, the aim of the task and the complexity of the whole activity. In addition to these types of exercises, two further types can also be used:

1. Cloze test: Basically this is a type of completion exercise with a regular slot missing. The procedure can be exploited in such a way that while content is controlled, various possibilities for text organisation can be left open. The learners, relieved of the responsibility of development of content, can concentrate their effort on the organisational nature of the text.

2. "Dehydrated" sentences: This type of exercise gives skeletal sentences with gaps and discontinuities, and the learners are required to reconstruct and restructure these sentences into continuous sequences (paragraphs or bigger segments), the content being related to the activity that the exercise is associated with. The value of this type of exercises rests in the provision of an excellent means for teaching use of connectivities (extension, variation, elaboration and/or enhancement) for organising text components.

3. Creative exercises

These can be of two groups. The first, like the exploitation exercises, comprise exercises related to an activity, but give the learner the opportunity to put down his own reactions and opinions

to its topic and content. The group may include a variety of exercises:

a. Reporting on an activity: This involves writing a brief account of the views in a passage or a film.

b. Taking minutes: This can be performed in group work while learners are engaged in an oral discussion or having a simulated session of some type (a committee meeting, for instance).

c. Responding to an argument, viewpoint or a "gossip".

These three types of exercises have to be structured in such a way that the written task can involve the use of connectives for introducing and integrating various statements. Particularly useful are adversative and orientative connectives for directing the arguments and providing the appropriate perspectives, and causal connectives for setting reasons and drawing conclusions.

The second group of creative exercises consists of more individualised and more creative activities. The learner is given the opportunity to write down his views or responses to a topic of his choice. It may be related to the classroom situation, for instance discussion based on pictures or cartoons, description of persons or objects, or an account of or reaction to a school event. Alternatively, the topic may lie outside the school. Newspaper stories, comments and views are particularly interesting. The learner, for instance, can be asked to write a letter to the editor or produce a short review or a response to a review. Again, the task has to be structured so that the learner is aware of the type of connectivities required for an efficient text composition.

10.4.5 Concluding Remarks

EFL teaching of writing is fundamentally an individualised operation. It has to cater for the learners' linguistic needs and respond to his personality and interest as well as his linguistic competence. To maintain interest and promote motivation, learners should see some purposes in the writing activity and, wherever possible, exert their own choices and preferences. They often need a clearly defined topic, an opening sentence, and a framework to get them started.

Further, writing practice needs to be structured in a way that secures optimal applicability. One of the essential requirements that have been mentioned is clear itemisation of tasks, so that each problematic aspect of connectivity can be catered for. This requires two preparatory operations: disintegrating the problem of text organisation into minute details for teaching purposes, and designing the appropriate practice material.

The process of itemisation should, however, not compromise the main aim behind the effort: familiarising learners with English conventions of textuality by acquainting them with patterns and strategies of various types of connectivities. Hence, ample integration is required so that the various exercises and teaching activities can, in their totality, contribute to competent text production.

Needless to say that considerable amount of planning, selection, and grading is dictated by feedback and results of error analysis.

This helps pinpointing with accuracy the more recurrent problems and gives an evaluation to the overall teaching effort.

10.5 Suggestions for Further Research

We will now venture a few final suggestions on possible future work that is either based on or inspired by this project. To systematise these suggestions, and due to the interdisciplinary nature of this study, we shall consider extensions of the project in four areas. The first involves further consideration of textuality, particularly within a contrastive perspective. The second assesses possibilities of developing computer-aided systems that can operate on the machine-readable corpora that we have assembled. The third involves lexicographical projects based on the various analyses performed on the corpora. And, the fourth lies within the realm of EFL pedagogy and concerns further investigation in problems of text production.

10.5.1 Text Analysis

In this study we have shed light on interlingual variations concerning two aspects of cohesion. One represents cohesion as a typology of semantic relations that obtain across various text components and help give the text a sense of tightness. The second aspect is related to the process of linearisation, that is the total operations whereby textual components are grouped and sequenced and textual salience is achieved and maintained. Examination has focused on the cohesive functions of connectives in achieving textual "tightness" and in assisting linearisation. The analysis can be extended in two types of projects:

1. One is devoted to a detailed corpus-based investigation of the nature of textual linearisation in English and Arabic. More specifically it examines the nature of large textual sequences such as the paragraph and paragraph cluster and investigates a) differences, and we presume their existence, in the way each language apperceives experience, and its effect on organising text world models; b) variations in the way conceptual configurations are grouped into "chunks", reflecting a paragraph, or other text constituent, division; in other words, variations in the conceptual structure of paragraphs and their contribution, through conceptual connectivities, to the coherence and informativity of the finished product, c) the grammatical/rhetorical means that create the differences in (a) and (b) above, particularly those preferential structures that map conceptual units into surface realisation (through a process of actualisation).

2. Another project is an extension of the study of textual cohesion initiated in this study to cover other cohesive means. This involves a contrastive account based on categorisation and quantification of each phenomenon. In this respect, we suggest a curtailment of the tendency to treat the cohesive devices as if they serve to refer to surface words and structures rather than to the conceptual-relational content underlying them. An alternative tendency is to focus on a diverse range of relations like inclusions among classes, superclasses, or metaclasses, part/whole, causality and proximity, particularly in the analysis of referential or lexical substitution or an extension of our present analysis of

connectives. A related examination may focus on the contribution that these means make towards processing efficiency and textual effectiveness. We are still far from having a satisfactory answer to how preferential choices of cohesive means can ease rather than burden processing, i.e. assist in processing the largest amounts with the smallest expenditure of resources.

3. A third project involves a comparison of the textual role of connectives across text types. The question of text typology is related to the consideration of text communicative functions. It is now axiomatic (refer to discussion in Chapter 3 of this study) that linguistic communication always appears in textual form. Since texts manifest different conditions of origin, different structures and are designed for different recipients and multifarious purposes, it follows that texts reflect a spectre of communicative functions. Texts with a comparable basic communicative function can be combined into a text type, such as instructional text (cf. Hartley 1985), business writing (from memos and letters to complex proposals, cf. Brown and Herndl 1986), etc. The investigation considers how connective functionality and patterning may vary in the way they support writing strategies, such as organising information from general to specific or vice versa, or introducing definitions, comments or illustration. The process can be analysed, quantified and compared, both within text types in one language or across two languages.

4. A fourth project is pertinent to Arabic and involves an empirical investigation of the variations, if any, in the utilisation of connectives in texts written by writers from

different parts of the Arab world. A tentative regional division may centre around four loci: Baghdad, Damascus-Beirut, Cairo and Rabat. The study can be extended to cover various modes of connectivities used as a basis for textual organisation.

5. A fifth project that is also specific to Arabic studies may investigate connectivity in classical Arabic writing and compare it to the results of our investigation. One important motivation is an examination of the type of organisational factors that have survived change and/or that still exerts considerable influence on Modern Standard Arabic.

10.5.2 Computerised Corpus Analysis

The presence of two comparable corpora in machine-readable form is an essential requisite for computerised linguistic analysis. At the time this project was conducted, and to my best knowledge, the Arabic corpus was the only fully-vocalised and statistically adequate form of Modern Standard Arabic available as a data-base for linguistic research. A number of follow-up computer-assisted projects can be initiated.

1. An attractive area of computer-assisted investigation concerning Arabic is the design of a system that achieves an automatic (or at least semi-automatic) tagging of Arabic (cf. discussion of types of tagging systems in Chapter 5, Vol.1). Two levels of tagging can be envisaged. One is a syntactic marking of the grammatical categories to which an item in the text belongs: grammatical function, case, mood, gender and number. The other is

more difficult and is still experimental at the present state of the art: introducing discourse tags that mark elements whose functions are typically bound to the communicative/rhetorical function of the text stretch where they occur. One example for such expressions is connectives, the identification of which cannot adequately be accounted for at the grammatical analysis level.

2. A related project involves the design of a system for an automatic lemmatisation of Arabic. Such a system is intended to perform an analysis of lexical patterns and identify the roots with a high degree of accuracy. Automatic generation and computation of roots and patterns have been the target of some research attempts (cf., for instance, Attalah 1973).

10.5.3 Lexicography

Again, the corpora that we have assembled can assist in carrying out a number of projects in the field of lexicography. We propose two :

1. The first is the preparation of an Arabic frequency dictionary of newspaper texts. This can be compiled on a similar basis to Allén's frequency dictionary of Swedish and Knowles' frequency dictionary of Polish. The word lists provided in the microfiche appendices at the back of this thesis can provide a basis, probably a strong component, for the dictionary. We envisage the dictionary to comprise the following parts:

a. An exposition of the way certain lexicographical problems have been treated in relation to Arabic. One such problem is the

important delineation of the concept of word as a unit of linguistic measurement (Appendix 1 in this work).

b. An alphabetical list of all vocabulary units appearing in the corpus together with their absolute and relative frequencies.

c. An alphabetical list of roots of the vocabulary units appearing in the units.

d. A detailed description of rank distribution listing all vocabulary items that occur within a rank. A statistical profile of each rank can then be given including such indices as: rank absolute and relative frequencies, absolute and relative accumulation by text, absolute and relative dissipation by text, absolute number and proportion of vocabulary, absolute and relative accumulation of vocabulary, absolute and relative dissipation of vocabulary, rank and individual item probabilities and repeat rates, longest, shortest and mean word lengths in the rank and rank entropy.

e. A frequency account of morphological patterns occurring in the corpus. These may include both verb and noun patterns.

Lists (c) and (e) assume the availability of sufficient computerised grammatical analysis that assist in the calculation of roots and patterns and of various grammatical categories.

2. A second project is based on the connective word lists and concordances produced in this study. It aims at producing a contrastive dictionary of connectives, or linking words, in English/Arabic discourse. Such a book constitutes a reference for the EFL Arab learner, but it may also be of considerable benefit to

the learner of Arabic as a foreign language. The functional categorisation of connectives furnishes an adequate basis for the organisation of the dictionary, while the numerous authentic excerpts that the concordances can provide serve to contextualise the textual patterning of the various connectives, thus strengthening its pedagogical applicability.⁽⁸⁾

10.5.3 EFL Writing Pedagogy

1. One immediate project that is a natural follow-up to this project is producing an EFL writing manual that is aimed at the Arab undergraduate majoring in English. The manual is expected to incorporate the suggestions that have been made in this study and caters for such factors as variability of previous linguistic experience and motivation.

2. Another involves introducing efficient schemes for evaluating EFL written composition on the basis of the results of this study. This involves assessing such aspects as text organisation and the extent to which a text meets the demands of English textuality. More specifically, we propose that a project in this area of investigation should attempt at answering questions such as: a) How can growth in writing be measured efficiently? In other words, what textual requirements does a teacher or an examiner look for in a text composition written at a particular stage of learning? Is the evaluation restricted to the written product? If it is, then what facets of the product are to be examined? Alternatively, can evaluation be extended to include the strategies or process which generated that product? If the answer is positive, then what

aspects of the process are to be measured? b) How can writing be evaluated over a period of time? c) How can the learners be involved in the evaluation procedure so that a maximal understanding can be secured of the textual strengths and weaknesses in a piece of writing?

10.6 A Final Note

At the conclusion of this study, we would like to stress the value of an interdisciplinary text-based type of contrastive studies that aim at exploring a) the various shared and unique procedures that establish connectivity of text components, b) the interdependence between the syntactic, semantic, pragmatic, and rhetorical constitution of the text, and c) the factors that give the text its functional unity and communicative value.

In this project, we have undertaken to show that the results of such studies make a substantial contribution to a better understanding of the interlingual variations of text actualisation and the divergences in the organisation of the finished product. This insight in textuality constitutes a prerequisite for the necessary ramification of pedagogical targets, particularly those related to the teaching/learning of written text production in a foreign language. Most important, in our view, is the gain that the applicability of the textual findings can bring to the didactic efficiency of the EFL teaching operation and the appropriateness, not to mention precision, of its methods.

(1) Some linguists have proposed other entities. Hinds (1979) puts forward the concept of the "segment": a smaller unit than the paragraph. Van Dijk (1982) proposes the "episode", a unit different from the paragraph in that it is a semantic unit, while the paragraph is its surface manifestation.

(2) One of the most prominent limitation of the orthographic paragraph is the indentation for each change of a speaker in a dialogue where the unity of dialogue paragraphs (in which, for instance, a question or a comment and its response constitute a unit) is obscured (cf. arguments and proposals in Longacre 1979).

(3) This principle has been the basis of instruction on paragraph development in writing manuals.

(4) The two roles of additive connectives of continuity in Arabic are dictated by strong textual pressures to maintain stability. The first role is manifested in the requirement of Arabic text organisation to signal continuity among text sequences. Despite the presence of various cohesive devices, text sequences, particularly sentences and paragraphs, remain discontinuous, i.e. disintegrated, unless continuity is resumed via additive connectives. English, on the other hand, assumes continuity without an urgent or necessary requirement for explicit signalling. This explains the high frequency of additive connectives of continuity in Arabic compared to English (see 8.2.1.2 and 8.2.2.2; see also 9.2.2.1). The second role, far more distinctive in Arabic than English, is indicative of the inherent inability of most Arabic connectives to sustain continuity among sequences. To support the cohesive role of these connectives, provision of continuity is made via the use of the connective "wa", which is positioned immediately before the supported connective. This arrangement is sufficient to secure textual continuity. Simultaneously, it attaches considerable importance to activities of extension for organising text and maintaining cohesion among its components.

(5) For instance, the so-called "cleft" construction (a dummy subject and predicate "it + be") delays communication momentarily in order to single out a particular item in the predicate, which would otherwise receive less attention.

(6) A number of EFL course books concentrate on developing this aspect. This is due to the strong tendency to teach the language sentence pattern by sentence pattern, with vocabulary being fitted in according to the situations used to illustrate the sentence patterns being presented. I quote, by way of example, some of the

various course books that I wrote in collaboration with Professor K. I. Al-Hamash and his team at IDELTI, Baghdad, Iraq, during the period 1977-1982; see, for instance, Al-Hamash and Al-Jubouri (1978, 1980, 1981) and Al-Hamash, Al-Jubouri and Al-Muttalibi (1978). In testing writing, lexico-grammatical components are among the main features to evaluate in students' compositions (cf. Al-Hamash, Al-Jubouri and Al-Hiti 1982).

(7) Due to limitation of space, and in order not to overload this chapter, considerations of goals of a writing programme are here overlooked.

(8) A dictionary of link words has recently been published (Ball 1986). Although this dictionary represents a useful contribution to EFL pedagogy, we have our reservations on the size and nature of its content and method of functional categorisation. We believe our study can provide the author of this book and the publishers with a useful basis for their future revision.

Appendix (1)

Reflections on the Word as a Unit of Linguistic Measurement

App. 1.1 Introduction

In determining the size of the corpora, an early tendency that asserted itself was to absolutise the identity of the word as a unit of quantitative linguistic description, and it was felt that it was wrong to suppress that tendency. English orthographic words, it was realised at the outset of the project manifest only partial correspondences in form and content⁽¹⁾ vis-a-vis Arabic orthographic words. Any attempt at a proper contrastive measurement will fall short of accuracy unless the identity of the word, both as token and as type, is defined with reasonable precision.

Sufficient motivation for resolving the problem of the word was provided by the fact that in Arabic orthography the connectives wa and fa, the two most frequent, are usually joined in writing with the following string of characters, making one single orthographic unit. The decomposition of such units, by disconnecting these two connectives, will give them an autonomous identity. But this procedure requires theoretical support, without which it might seem an implausibly clumsy procedure and would raise a number of theoretical as well as quantitative considerations. One such consideration is whether adequate justification is available for disconnecting a number of other strings that belong to various word classes. Another is how true is the image of the type/token ratio and other related formulae of measurement in depicting the

characteristics of each of the two corpora in general and of textual features such as connectivity in particular.

The crucial requirement for a clear definition of word prior to statistical word count cannot be overemphasised. Generally, in quantitative studies, it is imperative for a sound and competent analysis that the individuals, the populations and the characteristics observed are unequivocally delimited. This requirement may seem to make it appear hazardous to work with "words", since the word, though an ancient concept, especially in societies with a long literary tradition (to which both English and Arabic belong), and despite its familiarity as a unit of linguistic description, is still far from having secured an explicit definition (cf. Lyons 1968, Crystal 1971 and Robins 1964/1980). The lexical statistician and the compiler of a word frequency index, who, in the face of so disappointing a fact, persist in working with words as their individual observations, may find themselves compelled to use the orthographic form as their elementary unit of analysis, albeit their full awareness of possible adverse effects on the final results. However, in a matter of such fundamental importance as the procedure of word counting, the investigator, and in particular the one who is involved in the contrastive quantification of two languages, should not be content with the limits imposed by this shortcoming. What is required is a careful analysis of the linguistic evidence introduced for the description of word in the hope of systematising any subsequent investigation he carries out in this area.

Such linguistic evidence is not small in size. A vast amount of

research has indeed been devoted to analysing the concept of the word, a full review of which will certainly be outside the scope of this study.⁽²⁾ It suffices to say that linguists take different positions regarding this concept. Some deny the identity of the word. Seiler (1964), for instance, is of the opinion that all the troubles caused by the problem of the word originate in the fact that the word is regarded as a unit. To him "the word is not a unit but a constituent of a sentence or clause" (p.767). Bally⁽³⁾ rejects the concept of the word because in establishing it the standpoints of vocabulary, grammar (language system), phonology, and orthography are allegedly intermixed. Instead of the word he employs, on one hand, the concept of the "sememe", i.e., the mark for a lexical concept, and, on the other hand, the concept of syntactical molecule which includes both the sememe and the grammatical mark.

Some other linguists, however, do not reject the concept of the word, but restrict its use. Garvin (1964) regards the word as the properly defined distributional framework necessary for a precise determination of the distribution of morphemes. Bazell (1957)⁽⁴⁾ gives six criteria for the determination of a sound complex as a word: (1) insertion, (2) substitution, (3) sequence, (4) independence, (5) phonemic structure, (6) non-phonemic structure.⁽⁵⁾

Fully aware of the complexity of the concept of the word, and of the need for a definition to render the quantification procedures operationally feasible, we shall make it the aim of this part of

the study to, first, consider the word as the elementary unit of running text, discussing some procedural problems of quantification, and then examine and verify some of the existing criteria set for defining the word. We shall then adopt criteria capable of maximum application within the project so that a practical working definition is deduced that can embrace both English and Arabic. Later, our examination will be extended to include the word as a lexical unit in quantitative analysis, with the aim of deriving conclusions capable of serving as a procedural basis for the project.

App. 1.2. The Word as an Elementary Unit of Running Text:

The first task is to examine the word as an elementary unit of running text, or as a unit of (N), which here refers to the total occurrences of linguistic forms in a text, or to "tokens". In a considerable number of quantitative descriptions, the orthographic and typographic tradition of the language, or languages being described, have suggested a solution that has been generally recognised. However, the dispute that the identity of the orthographic word can create when employed as an elementary unit of running text has been expressed by a number of investigators in the field, and a demand for a definition has been verbalised (cf., for instance, Herdan 1964 and 1967, Muller 1963, Johnson 1976).

Muller (1963) maintains that if by orthographic unit is denoted a string of characters bounded by space or by punctuation marks, we can admit the equivalence

$$1 \text{ orthographic unit} = 1 \text{ word.}$$

However, this rule, like almost other rules of language, has its

exceptions, most notably in Arabic, where the orthographic unit can represent more than one component, each being a candidate for a definition of "word"; in other words:

1 orthographic unit = more than 1 word.

As referred to earlier, one may take the decision in instances like these to decompose such units into their constituents for the purpose of word counting. The consequence this procedure can lead to is diminishing the number of types, or vocabulary (V), and considerably enlarging the tokens, or occurrences (N). Such a decision, as has been mentioned before, should be guided by a competent delimitation of the identity of the word. But before this is pursued any further, a brief account of the orthographic word is to follow.

App. 1.2.1 Problems of the Orthographic Word

App. 1.2.1.1. The Arabic Orthographic Word

The development of the Arabic script is influenced by the phonological structure of the spoken language. In its beginnings, Barr (1976) states, it may be considered as "something like a graphemic representation of speech phonemes on a one-to-one basis (but with consonants only, plus some vowels)" (p. 99).⁽⁶⁾ These vowels are the long ones: a, u, i. Short vowels are indicated by diacritic markers placed above or below the consonantal skeleton.⁽⁷⁾

Like most alphabetic scripts, word boundaries in Arabic orthography are designated by the use of space. But in speech there is generally no systematic marking of word boundaries. This can be

manifested when a first time contact is made with a foreign language. Crystal (1971) argues that in such a situation the sentences may stand out fairly distinctly, "but inside these there is no way of telling, without further study, how many words a sentence consists of or where the boundaries between them go" (pp. 188-189). The utilisation of juncture as a boundary is rejected by a number of linguists (Crystal 1971, Lyons 1968, Southworth and Daswani 1974, Henderson 1982). Henderson (1982) maintains that we owe to the spectrogram the discovery that the segmentation of speech stream into words is a perceptual achievement rather than a physical one.

This problem is further illustrated by the highly technical discussion of what constitutes a "word" to be found in Chomsky and Halle (1968, pp. 366-371). In this monumental work, Chomsky and Halle specifically state the assumption that word boundaries are systematically indicated by phonetic effects. In their model, words are segmented by a complex process of rules. The convention that is postulated for word boundary is stated as follows:

'The boundary # is automatically inserted at the beginning and end of every string dominated by a major category, i.e., by one of the lexical categories "noun", "verb", "adjective", or by a category such as "sentence", "noun" phrase", "verbs phrase", which dominate a lexical category.' (p.366).

In addition to this convention, Chomsky and Halle assert the existence of language-specific rules governing the presence of #. Conceivably, they argue, there may be rules that introduce # in various positions not specified by this convention, although they admit, they know of no clear examples of this; however, they

maintain there are rules that delete # in various positions.

One property of this analysis that concerns us here is that the "words" it yields sometimes consist of phrases in which a preposition and/or article are combined with the following noun. To quote an example from English, the sentence "the book was in an unlikely place" is considered by Chomsky and Halle to be composed of three words: the book, was in an unlikely, place. The authors conclude that "the orthographic conventions for Hebrew and Arabic... are consistent with the phonetics in not separating prepositions or articles from the following word" (p. 368).

Before we go on, two comments on the last statement are mandatory. First, only a few prepositions in Arabic are not orthographically separated from the following word, and these include bi, li and ka. Other prepositions are orthographically separate forms, e.g. <alā, fī, 'ilā, etc. It seems that an additional factor lies in the fact that li, bi, ka as well as the connectives wa, fa and the future particle sa are each made up of 1C + 1 short V. In writing, these stand out as one consonant letter, with a diacritic marker above (fatha), or below (kasra) in pointed (vocalised) writing, or without any diacritic in the normal unpointed (unvocalised) writing. A speedier and more efficient, and therefore more convenient, manner of writing that has historically become conventional, is to merge such words with the next.⁽⁹⁾

The other comment concerns certain Arabic pronouns: the type labelled "connected" by grammarians, as opposed to "separate" (which is an orthographic arrangement). These are joined with the

preceding word, e.g. katab tu and kitab i (written as katabtu and kitabī).⁽¹⁰⁾ Rules are required to formulate such a phenomenon.

If we are to accept Chomsky and Halle's statement concerning Arabic, we must introduce some modification. The Arabic script, in attempting to segment speech stream into words, groups certain function words along with following or preceding associated content words, so that the final product is one single orthographic word. Accordingly the expressions bi ḥaḳībati hi [in his bag] and qābal tu hā [I met her] are written as one single unit each: biḥaḳībatihi and qābaltuhā.

Chomsky and Halle's model as applied to Arabic requires extensive empirical investigation to establish the nature, extent and limitation of the rules involved. In our present state of knowledge, caution dictates the assumption that there exist no wholly reliable phonetic features to signal word boundaries and therefore the Arabic orthographic representation of such boundaries falls short of accuracy. Until further research is achieved in this area that ushers to a different direction, which can result in a different conclusion, we shall reject the Arabic orthographic unit as representing the concept of word.

App. 1.2.1.2. The English Orthographic Word

In English orthography, the equivalence of 1 orthographic unit = 1 word, though causes far less problem than in Arabic, can itself in cases manifest a considerable degree of difficulty. One such case is the presence within the orthographic unit of one of two

characters which are not alphabetic and which can create sufficient doubt as to blur the identity of word. These two characters are the apostrophe and the hyphen. Another familiar case is where a unit contains numerical and alphabetic characters. A third case, as will be shown below, concerns acronyms and abbreviations.

1. The apostrophe: With regard to the distinction in which we are interested, this character must be considered as ambiguous. On the one hand its integration with the alphabetical characters means to make it part of the orthographic unit, thus making they're and boy's one unit each. On the other hand, regarding it as separating two words means decomposing such units as o'clock and don't into two words each: o and clock, don and t.

2. The hyphen: The hyphen is just as ambiguous in this respect as the apostrophe. Generally its main function is "to separate the prefix of a newly (or temporarily) formed word" (Quirk et al. 1972, p. 1057). Often it is used in creating "compound" words by joining two or more bases, such as "tax-free" and "colour-blindness" where the compound is considered to form one concept. However, orthographically, compounds can be written as either solid (e.g. bloodtest) or hyphenated (e.g. tax-free). In certain instances a compound can be written either solid or hyphenated (e.g. flowerpot, flower-pot). Or, a compound noun can be written as one hyphenated word or left as two separate words (e.g. letter-writer, letter writer, air-brake, air brake). Quirk et al. (1972) comment on this point stating that there are "no safe rules-of-thumb that will help in the choice between these three possibilities" (p. 1019, notice their interesting hyphenation of "rules-of-thumb").

3. Strings with numerical and alphabetic characters: These are written as one string, though one can argue that they stand for more than one word. If "22bn", for example, is read aloud: "twenty-two billion pounds", it renders three (or four?) words. A minor problem in computing is the numerical string with ',' such as 123,456, or the decimal point '.' as in 12.34. Since both ',' and '.' coincide with the comma and full stop (two non-alphabetical symbols that can mark word boundary), it has the effect in computer processing of dividing one numerical string into two words.⁽¹¹⁾

4. Acronyms and initials: These pose two types of problems with regard to the distinction in which we are concerned:

a) Does the orthographic unit USA or OPEC stand for one or more words?

b) Occasionally acronyms in printed material are written in such a manner that spaces are introduced to separate the characters that make up the acronym: 'USA or OPEC.'⁽¹²⁾ These spaces divide them into several words and has the effect of enlarging both (N) and (V) within the corpus.

Problems that the orthographic unit causes have to be settled before a word count is started. This can only be done by delimiting the concept of the word. This can be arrived at by adopting a set of accepted criteria. But first an account is given and examination made of the criteria that are available in the literature for the determination of a unit as "word".

App. 1.2.2 Criteria of the Word

Despite the pessimism of some linguists about the possibility of defining "word", a number of criteria have been proposed.

App. 1.2.2.1 The Semantic Criterion

According to this criterion, a word is determined semantically by its meaning. Sapir's definition "the word is one of the smallest completely satisfying bits of isolated 'meaning' into which the sentence resolves itself" (1921 p. 35) and Palmer's "the smallest speech unit (=constantly recurring sound pattern) capable of functioning as a complete utterance", can be considered as based on a purely semantic criterion. Bloomfield (1933) employs the same criterion when he defines the word as the smallest expression of meaning.

Unfortunately, the exclusive employment of this criterion has given rise to at least two sets of difficulties. One concerns the group of words labelled "synsemantic" as opposed to "autosemantic".⁽¹³⁾ According to this criterion "synsemantics" (i.e. words which by themselves cannot realise the sentence, or can do so in very rare cases only, cf. the English my, your, the, a, or the Arabic bi, hum) should not be regarded as genuine words. Otto⁽¹⁴⁾ (1943) rejects this opinion as unrealistic, and sets forth to ascribe to the word a double meaning: Begriffsbedeutung and Beziehungsbedeutung. Even prepositions and conjunctions, he maintains, have this double meaning. Marty's conception of the word (1925 p. 40) does not give up the semantic criteria, but his cautious formulation leaves the door open to synsemantics too.

However, Kramsky (1969) points out that Marty's definition lacks a more clear-cut stress of the independence of the word in contrast to the morpheme, which constitutes the second type of difficulties with the semantic criterion.

To avoid the overlap, created by the semantic criterion, between the morpheme and the word, Martinet (1962) coined the term "moneme". Roughly, monemes are what is usually called morphemes; thus a word like reading would consist of two monemes read and ing. Martinet warns that the moneme should not be considered as a mere technical substitute for "word", and, in supporting this view he uses the criterion of separability, the one which will be outlined next.

App. 1.2.2.2 The Criterion of Inseparability

This criterion, referred to as "non-interruptibility" by Pike and Pike (1977), is formulated by Vachek⁽¹⁵⁾ in the following way: "In so far as a certain sound section has the validity of a coherent word whole, it is not possible for any part of it to separate from its other parts".

This characteristic is often referred to as "internal stability" (Crystal 1971, Robins 1964/1980) or "cohesion" (Kramsky 1969). Words interrupt; they are not interruptible. They do not permit internal rearrangement of their constituent parts, nor the insertion of comparable and virtually unlimited further stretches of utterance.

Accordingly, a number of tests have been set up to test inseparability.⁽¹⁶⁾

1. Replacement: This refers to the paradigmatic capacity of the word. A unit is to be called a word if it can be replaced. Thus in English the article a in a book is a word since it can be replaced by the or my, each is also a word. By the same token, in Arabic li in li miṣra must be a word since it can be replaced by fī.

2. Displacement: This refers to word identification in terms of "minimal unit of positional mobility" (Crystal 1971, p.190) and means that a linguistic unit is a word when it is capable of being permuted within the sentence.⁽¹⁷⁾ This test has been approached with caution, and Kramsky (1969) agrees with Hořejší (1957) in not regarding this test as reliable, since displaceability is limited by the combinatorial capacity of the words and their variable degree of mobility (some words like the are positionally fixed). Furthermore the displacement of words without any distortion of meaning is not permitted in English where the morphological functions of the words are usually expressed by word order, and not by word endings.

3. Insertion: According to this test, two units are words if it is possible for another form (or forms) to be inserted between them. Kramsky (1969) qualifies this test by adding a condition according to which the separating element must be an independent word into which it itself could again be separated. However, one problem that this test can give rise to, even if Kramsky's condition is fulfilled, is regarding the genitive 's as a word, since in the King's an independent form can be inserted: the King Arthur's. Hořejší accepts that the conception the the genitive 's is an independent word. Kramsky, however, rejects it, maintaining it is

at variance with another criterion which he advocates: the criterion of isolatedness.

App. 1.2.2.3 The Criterion of Isolatedness

The origin of this criterion, which is discussed in a number of works (cf. Kramsky 1969, Falk 1978, Robins 1964/1980, Crystal 1971) and adopted as a major criterion by Elson and Pickett (1968) can be traced back to Bloomfield's "minimal free forms", which are further defined as "the smallest items which are spoken by themselves in isolation". (Bloomfield 1935, p.160)

This view is shared by Polivanov (1936).⁽¹⁸⁾ In resolving the question of what can be considered a word, he maintains, it is possible to apply the criterion of the isolatedness of the word. The word, he adds, is a potential minimum sentence, i.e. a segment of speech which can be isolated and which can be uttered as a single sentence component (e.g. in conversation when given as a partial repetition of statements, questions and answers). Accordingly Polivanov considers a sequence such as "je te le dis, je te l'ai dit, je ne dis pas, je ne te le dis pas" as particular words composed of elements which cannot exist in isolation.⁽¹⁹⁾

This definition is questioned by Crystal (1971) and by Kramsky (1969), who, along with Togeby (1949), pose the question whether English forms such as the, a, is, and are ever spoken alone. Some other linguists, such as Bloch and Trager (1942), who have employed this criterion, have supported it with the semantic criterion.

1.2.4 The Grammatical Criterion

According to this criterion, a linguistic form is regarded a word if it can be allocated a grammatical "class" or "category" in the syntax of the language. In the traditional Arabic grammar, words are classified primarily on the basis of their internal structure thus yielding a tripartite division consisting of nouns (generally with such affixes as indicating number, case, etc.), verbs (with such affixes as indicating tense, aspect, etc.) and particles (uninflected words).

One of the arguments put forward in favour of this division in Arabic is that within a structural perspective a word is either capable of functioning as a subject or predicator (the noun) or only predicator (verb). If it cannot function as either it is a particle. However, most Arab grammarians employ a semantic criterion as well to consolidate the grammatical one in their specification of the word (cf. for instance Ibn Hušām "Šuḡūr Al-Dahab", Ibn ʿaql "alfiyyat Ibn mālik", Al-Rādiy "Šarḥ Al-Kāfiya").

This criterion has also been proposed by some structural linguists (cf. Southworth and Daswani 1974) and accordingly the word is defined as "any sequence of morphemes which never cross the IC boundaries" (ibid p.135). A weakness in this definition can be located in certain cases where a sequence can be accepted as two words. For example, in a form such as "indescribable", the IC cut can create two constituents in/describable, thus resulting in two words.

The adoption of the grammatical criterion, among others, is

succinctly expressed by Pike and Pike (1977) in their tagmemic grammatical analysis. The criterion, labelled as "slot-role proportion" is convincingly stated as follows:

"When there is a filler class of elements, one of which is a word by [criterion of isolatedness], and another of which is not, the second may be called a word by analogy with the first, if the two have the same proportional slot-role relation to the context." (p.113; their underlining).

Thus, according to this criterion, Pike and Pike add, if in reply to the question "what bit you?" one can obtain either "That wild cat", or "That", or "A wild cat", then although one does not obtain A as an answer, the morpheme a may be considered to be a word because it is in the same filler class with that.

Having outlined some of the main criteria for the definition of the word, our conclusion at this stage, which will be elaborated below, is that there is no one single fail-safe criterion capable of maximum applicability, especially in a project of a contrastive nature. The judgements arrived at by two investigators may differ in weighing the importance of each set of criteria for any one language. This can yield, as a final product, different descriptions. The size of the difference can be minimised by the openness and flexibility of the approach adopted. This will be elucidated in the next section.

App. 1.2.3 The Word in the Two Corpora

The heavy reliance on one particular set of criteria in preference to others is mainly responsible for the confusion the

term "word" causes in discussions about language. Linguists recognise four types of words: phonological, orthographical, lexical and grammatical. It is true that there is a high degree of co-extensiveness among the four: many formal items, like "pen" may at one and the same time be an orthographic word, a grammatical word, lexical item, and a separate phonological entity. But even in such cases, it is important to recognise the four different processes of identification that are involved (cf. Halliday, McIntosh and Stevens 1964, p.36).

It is our contention, and in this respect we are in agreement with Robins (1964/1980), that words, like many other analytic entities, are the products of several different, though related, criteria. Within a language, the category of words comprise three types of members: (a) nuclear members of the category, to which all the criteria apply; (b) more peripheral or marginal members to which some criteria only apply; (c) doubtful members in which the criteria may conflict, in this case different conclusions can be reached by the different weighting of the conflicting criteria.

Priority within the order of the criteria has to be established. In reviewing the criteria, some are problematic in their application and accordingly will occupy a lower place in the order of priority. Others are more distinctive in demarcatively shaping the entity of the word, and consequently are in the top of the order. It is ancillary to our treatment, and therefore will not be attempted, to examine the way a totally different order of priorities from the one adopted here can contribute to resolving the problem under discussion.

In attempting to define the word, the first criterion we shall apply is the semantic one: a linguistic form is a word if it conveys meaning. We agree that the term "meaning" is itself complex and in many cases ambiguous. However, as it is employed here, the term has a general coverage, the product of a number of semantic approaches (referential theory, ideational theory, behavioural theory and the theory of use).⁽²⁰⁾

This semantic criterion covers not only content words (autosemantics), but it is sufficiently flexible to include function words (synsemantics). In this, we are in agreement with Marty's formulation.

In cases where the independence of the word is questionable according to this criterion, we resort to the next two criteria to help resolve this problem.

The second criterion to be adopted here is the grammatical one. A unit is a word if it is capable of realising two potent series of affiliations: syntagmatic and paradigmatic. The relations are here viewed from a grammatical angle (semantic relations have already been mentioned). In this respect, paradigmatic links impose that a word class (traditionally, parts of speech) is assigned to words. Syntagmatic relations regulate the manner in which these units enter not only as components of the phrase, sentence or text, but also in connection with other words.

The third criterion is internal stability (as used by Crystal 1971 and Robins 1964/1980). Units that have a fixed internal

structure, in the sense that bits which constitute them cannot be rearranged in any way, nor can they be separated or replaced by other units. This criterion, stated in this rigid manner, requires some flexibility. Words in different languages, and even within one language, can exhibit a variable degree of closeness of the components that enter into their structure. The highest degree of cohesion is realised when no element can be inserted between the components of the word. However, Kramsky (1969) distinguishes different degrees of closeness (or "cohesion") according to a varying degree of obstruction, and concludes that languages in this respect behave differently.

These three criteria, we believe, should suffice to delineate the identity of the word in our project. In the case when these three criteria are co-extensive with the orthographical features, a maximal unity of structure is brought into prominence and the identity of the word can totally coherently be delimited. This makes up the nuclear members of word mentioned above. Where these criteria are not maximally congruent with the orthographical features, peripheral or marginal members are identified.

However, in the case of dispute or doubt, the fourth criterion to consider is that of the lexicon. A unit is a word if it is found in the lexicon of the language as an independent unit. In the settlement of disputed units in Arabic, Wehr's Dictionary and Al munjid are consulted. We are aware of the fact that different lexicographers may organise the lexicon differently, depending on a host of linguistic and extra-linguistic factors; however, the

extensiveness of these two works and the intensity of their use give us sufficient confidence to treat them as highly authoritative.

A working definition of the word as an elementary unit of running text which is to be adopted in this project for both English and Arabic, and which can now be deduced, is compatible with the one proposed by Kramsky (1969 p.67). The word is defined as the smallest independent⁽²¹⁾ unit of language referring to a certain extralinguistic reality or to a relation of such realities and characterised by its capacity to express syntagmatic and paradigmatic relations, either actually (as an independent component of the text) or potentially (as a unit of the lexicon).

App. 1.2.4 Application to Arabic Texts

The impact that our conception of the word has on Arabic text is immense. The following units, long established orthographically as part of the next or preceding unit, can now each be given the status of a word. Accordingly they have been disconnected and a space is created to distinguish them from the words to which they are traditionally forced to amalgamate. The units, or words, are:

1. All connected pronouns: tu, na, ta, ti, tumā, tunna, hu (hi), ha, humā, hum, hunna, nā, ī, āni, ūna, īna, ā, ū (often followed by unpronounced 'alif and is represented in the corpus as UL), ka, ki, kumā, kum.

2. The prepositions: li (la, l), bi, ka.

3. The particles: li, sa.

4. The connectives: wa, fa.
5. The question-word, 'a.
6. The definite article Al.

App. 1.2.5 Application in English Texts

The working definition that has been formulated helps resolve a number of the problems associated with the English orthographic word. The genitive 's is one such problem (as has been mentioned above). While 's can satisfy a semantic criterion, it violates the other criteria set for the definition of the word in this project. It follows that the genitive 's is regarded as a bound morpheme, and not as an independent word.

Strings that contain the apostrophe ' such as don't and they've are considered independent words, since contraction is often related to stylistic considerations, such as degree of formality that characterises the text as a communicative message.

The compound word (whether hyphenated in its orthographic realisation or merged) is regarded as a single word. This conception is regarded compatible with the criteria and the definition outlined above, despite its apparent incongruency with our awareness that a number of compound word (hyphenated or merged) can be written as two words by some writers. Since this orthographic difference can often be ascribed to personal stylistic choice, it has been accepted that such a choice is to be respected and maintained. Our intuitive feeling in this respect is that this procedure will not practise a drastically adverse effect on the

statistical computation of the corpus as a mass.

Strings that contain numerical and alphabetic characteristics are considered similar to compound words. Accordingly a string such as " 22bn" is regarded as a single word. The decimal point mark '.' is considered as a padding character and therefore a string such as 12.34 is considered a single word. The difficulty created by the presence of "," in numerical strings such as 123,456 has been overcome by deleting ",", leaving the string looking like this 123456. This procedure, despite its apparent clumsiness, has proved conveniently serviceable, and has indeed restricted potential errors in processing.

Acronyms have been reduced to a single string. Spaces or dots, have been deleted to render the string a single word. Thus U S A or U.S.A. is modified into USA. The effect is a more unified entry in the word lists and concordances that are later produced. Statistically, it reduces the number of types, or V in the corpus and offers a better descriptive image of its quantitative properties.

App. 1.3 The Word as Elementary Lexical Unit

The aim of the previous section was to delimit the identity of the word as an elementary unit of running text in the two corpora. The total of these units, or words, represents (N): the number of tokens, or word occurrences (cf. Williams' (1970) term "usages"). This section considers the word as a unit of (V), the vocabulary or total number of types (cf. Williams' term "uses").

In arriving at (V), the common practice is to re-group the occurrences for each lexical unit (i.e. each token) which are dispersed all over the text. The traditional basis for this re-grouping is the orthographic form, and accordingly the rule in the majority of cases is:

1 word = 1 orthographic form

This rule, however, is not without its exceptions, some of which, especially in the case of Arabic, are of a more serious character than the ones encountered when the units of (N) were discussed in the last section. The uncertainties that these exceptions have created have long been familiar to lexicographers of European languages (English in particular; for a good treatment of traditional Arabic lexicography see Haywood 1969). Generally speaking, lexicographers have at their disposal means of counteracting these uncertainties. Since they are not directly interested in frequencies, nor do they have the task of assessing the use of lexical elements quantitatively, they can afford to leave such uncertainties in suspense. However, the lexical statistician and the compiler of a frequency index must, at a given moment in their work, arrive at a decision that would remove all doubts. Such a decision, which will influence the size of (V) and affect the V/N relationship, is usually based on the options that are exercised in every case under examination.

The extent and the total of these options are largely governed by the purposes of the investigation. Herdan (1966) makes a

distinction between two types of investigations: investigation of language as a coding system and investigation of language for teaching (and other related) purposes. The first type aims to study the system of vocabulary in its properties as a coding system; hence, Herdan argues, we are only interested in the form of the words. The second type of investigation, on the other hand, aims at exhibiting parts of the vocabulary system for teaching purposes, hence, it is advisable and helpful if the frequency of vocabulary items is differentiated according to the meaning of words. This point will be taken up later.

To assess the type and the extent of options, it was considered useful to examine the ways some statistically oriented linguists and compilers of frequency indices have treated the concept of the word as a unit of (V).

App. 1.3.1 The Word in Some Frequency Lists

The words that will be examined include the frequency lists of Brown corpus (Kučera and Francis 1967, Francis and Kučera 1982), the LOB corpus (Johansson and Hofland 1982), Arabic prose (Landau 1954), Polish journalistic texts (Knowles 1981).

App. 1.3.1.1 The Word in the Brown Corpus:

Since its construction in machine-readable form, the Brown corpus has been analysed from various angles. Various word frequency analyses were produced, some are published in Kučera and Francis (1967) while others are in their more recent Francis and Kučera (1982). In the analyses of Kučera and Francis (1967) the "word" as

a unit of (N) is the orthographic word, simply defined as a continuous string of letters, numerals, punctuation marks, and other graphemic symbols, uninterrupted by space. The word as a unit of (V) in the analyses is simply defined as a set of identical individual words. The authors admit that, while basing the lists uncompromisingly on the graphic word as a unit restricts the usefulness of the counts, especially for stylistic analysis, it is hard to see how any other procedure short of a completely "semantic" count like that of Lorge is possible.

In their 1982 work (pp.3-4), Francis and Kučera differentiate a number of concepts for word. A graphic word is defined as "a string of contiguous alphanumeric characters with space on either side; may include hyphens and apostrophes but no other punctuation marks". A "grammatical word" in their work is a graphic word with its tag (which is a string of capital letters and/or symbols indicating the grammatical category or class to which a graphic token is assigned). A lexical word is defined as "one or more grammatical words forming a lexical unit - i.e. filling a single grammatical position and having a unitary meaning". A "compound word" is a lexical one analysable into constituents which themselves constitute lexical words. This is distinguished from a "merged word" which is defined as a graphic word made up of two lexical words, one or both of which are rendered in a reduced graphic form to indicate phonological reduction in speech (e.g. he'll). A "pseudo-word" is a term used to refer to "a graphic word not recognisable as a lexical word, usually because it is a hyphenated compound, particularly with compound proper names. These types of

words are differentiated from the "lemma" defined as "a set of grammatical words having the same stem and/or meaning and belonging to the same major word class, differing only in inflection and/or spelling".

Of these various conceptions of word, the units that constitute the data for the distribution and frequency are the grammatical words (ibid, p.16). Each one, including spelling variants, is considered a separate type and is therefore separately counted. In addition, subtotals are given for each inflectional and spelling variant and totals for each lemma. Since the purpose of the word counts is intended to be multi-fold, and therefore the lists are expected to profit a wide sector of investigators and users, the image that is proffered for each entry is comprehensive enough.⁽²²⁾

App. 1.3.1.2 The Word in the LOB Corpus

Similar to the Brown word-frequency lists, the LOB word-frequency lists treat the orthographic word as a unit of (N): a sequence of alphanumeric characters surrounded by spaces (Johansson and Hofland 1982). As a unit of (V), the word is, again similar to the Brown lists, a set of identical individual words.

App. 1.3.1.3 The Frequency Lists of Modern Arabic Prose

In determining the size of (N), Landau (1959) uses the orthographic form of the word as his elementary unit. However, in defining the word as a unit of (V) he employs criteria based on the work that was previously carried out by Brill and his assistants (1940). Both projects are carried out manually, a task, though

laborious and tedious to perform, has the admirable effect of including some type of lemmatisation and separation of homographs.

However, some of the decisions concerning the identity of the unit of (V) are questionable. Such decisions are discussed below.

1. Each verb is listed under its stem, and in the masculine third person singular perfect. This procedure, while compatible to some extent with the traditional method of arranging the Arabic lexicon, is, as employed by Landau, not itself a full lemmatisation of verbs. Additionally, it has the disadvantage of being uninformative regarding the frequencies of the forms of the imperfect.

2. Adjectives are listed in the masculine singular, except in cases where "the feminine is quite different". Landau does not explain whether the "difference" is orthographic, semantic or grammatical. In implementing this decision, feminine adjectives such as "kabīra" [feminine, "big"] are combined with masculine adjectives such as "kabīr" [masculine, "big"], while such feminine adjectives as "bayḍā'" are listed as a separate entry from their masculine counterpart such as "'abyaḍ". On the other hand, adjectives in the superlative degree, as a rule, are not treated as separate entries and therefore "'akbar" [masculine, "bigger, biggest"] is still combined with "kabīr". This procedure, inconsistent and unsystematic as it is, has not been linguistically, nor statistically, justified.

3. Broken plurals are listed separately from their singulars, e.g. "rijāl" [men] is a different entry from "rajul" [a man]. But

sound plurals (masculine or feminine) are combined with their singulars, that is the frequency of the singular represents not only occurrences of the singular form, but includes all occurrences of its plural as well. As no justification is offered, the procedure is here considered methodologically unsystematic, liable to considerable potential statistical inaccuracies.

4. While prepositions such as "fī" and "<alā" are included in the frequency list, the two prepositions "bi" and "li" are excluded except when they are amalgamated with "connected" pronouns. This procedure, both grammatically and semantically unjustifiable, seems to be influenced by the orthographic form of the word and, hence, can cause sizeable inaccuracy in any computation where (V) and (N) are involved, or where a statistical account of prepositions is required.

5. Similar to the procedure in 4 above, the connective "wa" and "fa" are considered as parts of the word they are orthographically attracted to and are therefore ignored in the frequency listing, thus rendering the lists inefficient for a quantitative account of connectives such as ours.

In brief, the shortcomings in Landau's frequency lists stem from his lack of a clear definition of the word. There are discrepancies in the criteria he sets for the recognition of the word that adversely affect the calculation of both (V) and (N) in his corpus. For instance, while he adopts a semantic criterion in his endeavour to separate homographs, he dismisses it altogether in some other instances and employs instead an uncompromising orthographic

criterion. His grammatical criterion, as in his treatment of the masculine/feminine distinction or in resolving the dichotomy of sound/broken plurals, conflict to an unfavourable extent with his other, semantic and orthographic, criteria, and is far from being rigorously applied. Perhaps, as a compromise, he should have prepared two types of frequency lists, each satisfying a different set of criteria, in order to minimise the contradictions that exist in his present one: the first could have been constructed on the basis of the orthographic word, while the other on rigorous semantico-grammatical considerations. This is what Kucera and Francis did in their lists of 1967 and 1982.

App. 1.3.1.4 The Frequency Lists of Polish Journalistic Texts

The frequency lists of Polish are based on a corpus made up of a selection of some forty-two articles from Polish journals (Knowles 1982, vol.1, p.i). The corpus is 46,000 tokens (N) in size which contains 6000 types (V). Since the aim of the lists is to produce a word-frequency dictionary for Polish journalistic texts, considerable "processing" of the original lexical data is necessitated. The two chief processes involved were lemmatising and editing. Accordingly, the shape and size of the word as a unit of (V) has been substantially affected. Some of the decisions that have been taken are:

1. Verbs are reduced to the imperfect infinitive (wherever this was possible).
2. Adverbs are reduced to the appropriate adjective, if it

exists.

3. Adjectives in the comparative or superlative are reduced to the positive degree.

4. All inflected words are reduced to the traditional "dictionary look-up forms".

These decision are taken with the full awareness that what is aimed at is a "dictionary", and therefore a lemmatised rather than unlemmatised version of the word lists are required as potential entries. This explains the withdrawal of an original partially lemmatised version of the lists created earlier (cf. Knowles 1972), so that the present version can replace it.

There is one comment to make about the present version of the dictionary. The lists included are the product of an investigation for teaching and for related reference purposes. Indeed another earlier product is the derivation of a "core" vocabulary of Polish for Russianists, a list comprising items in Polish which are either not familiar in appearance to or deducible from their Russian counterparts or which are worthy of comment in some other respect (Knowles 1972). Given the aims and direction of this investigation one can justify the compact nature of the Polish corpus and the seemingly arbitrary decisions taken in the specification of the entries. If however, an investigation aims exclusively to study the properties of language as a coding system, one would expect a larger corpus of text than the present one and more comprehensive information regarding the inflectional or spelling variants as well as a short appendix to indicate information (both linguistic and

statistical) on those elements that, for pragmatic and other methodological considerations, have been dropped out of the count.

App. 1.3.2 The Word as a Lexical Unit in the Project

The survey above has been intended to assist and direct our efforts to contrive a set of rules that will govern the judgement that a particular form is a unit of (V). The rule that has been mentioned above, i.e.,

1 word = 1 orthographical form

suffers from numerous exceptions that, especially in Arabic, tend to obscure the concept of the word. The exceptions, following Muller (1963, 1972), can be divided into two broad categories. These will be discussed in detail below.

A. One word = several forms

It is often the case, and in particular with a highly deflectional language such as Arabic, that a word can, owing to conjugation, inflection or deflection, have more than one form. The semantic unity of such different forms is beyond doubt, and indeed in the process of lemmatisation, the various grammatical modifications of a particular stem (lexeme) are deleted and often only one form is preserved to represent the word. However, even then, criteria have to be carefully designed to govern what forms to be lemmatised or not.

In both English and Arabic, morphological modifications that can affect the form of the word can be classified into two broad types

according to the nature of the cause involved: modifications due to various grammatical categories and modifications due to juxtaposition. These are outlined below:

1. Modification due to grammatical categories

- a. Person: Modifications due to person affect both languages and are applicable to pronouns and verbs. Both English and Arabic pronominal words are organised into a system of three persons, referring to the speaker (first person), the addressee (second person) and the person referred to but not addressed (third person). Moreover, different forms for each person category are found depending on cross-category intersection with gender, number and case (see below). The category of person in both languages require concord with the imperfect verb, usually realised as a marker (e.g. the use of "s/es" with the English present simple following a third person singular or the use of "'a, ta, ya, na" as the initial part of the imperfect in Arabic denoting first (singular), second (singular, masculine) or third (singular, feminine), third (singular, masculine) and first (plural) respectively.

- b. Tense: The category of tense imposes variations in the form of the "base" verb (the event). The semantic ranges covered by the forms are on two axes: time relations and aspectual differences in completion, incompleteness, continuation and momentariness (Robins 1964/1980 p.208). The system of tense is realised by the logical structure of the verb (or, more accurately, verb groups, cf. Halliday 1985, p.177). According to Halliday (1985), the realisation of tense is divisible into two

types: primary and secondary. The primary tense is that functioning as head in the verbal group and indicates whether the verbal group is present, past or future, relative to the speech event. The secondary tenses are modifying elements that express present, past and future, relative to the time selected in the previous tense. In Arabic the temporal forms are two: perfect (al-madi) and imperfect (al-mudari<), and therefore they correspond to the primary tenses. The future can be realised by the juxtaposition of the particle "sa" (positive polarity) or "lan" (negative polarity) to the imperfect. Secondary tenses and relevant aspectual markers can be contextually indicated and thus few grammatical (or morphological) paradigms associated with time distinctions are necessary.

c. Mood: We use this term to refer to forms in the conjugation of verbs as they occur in indicative or non-indicative constructions. It is the same use employed by Zandvoort (1962) pp.64, 86-9), Strang (1962/1971, pp.152-4) and Leech (1971, pp.106-8) in their description of the English verb (cf. Palmer 1979, pp.3-4 and Halliday 1985b pp.40-1 for a different treatment of verb mood). In this sense, English verbs are either indicative or subjunctive, the decisive marker is the presence or absence of -s from a present verb whose subject is third person singular number. This sense, Palmer (op. cit.) argues, is narrow and questionable, and the contrast between subjunctive and indicative moods has largely disappeared from present-day English grammar (Leech op.cit.) though the distinction of meaning which the subjunctive and indicative used to express is still important within the language.⁽²³⁾

In Arabic, however, mood is essential for the description of the syntax of Arabic words. It is a syntactic category that can inflict a number of morphological changes within the structure of the imperfect verb, thus causing variable modification in form. There are three moods that characterise the Arabic verb: indicative [raf<], subjunctive [nasb] and jussive [jazm], though Wright (1962, vol.1) adds two others: imperative and energetic.⁽²⁴⁾

In general, strong imperfect verbs (as opposed to infirm ones) carry a final -u (damma) as a marker of indicative mood and a final -a (fatha) as a marker of subjunctive. In the jussive mood an imperfect strong verb is characterised by the absence of a marker. Thus "yaktubu" is indicative as opposed to "yaktuba", subjunctive, and "yaktub" (jussive). When the strong imperfect verb precedes the "connected" pronouns "āni" (dual second/third person), "ūna" (plural, masculine, second/third person), ina (singular, feminine, second person), the verb itself loses its final mood-marker vowel. Indicative mood is then marked by the presence of the ni or na in the following connected pronoun, while the subjunctive and jussive are marked by their absence, thus changing the pronouns into ā, ū and ī as in "yaktub āni" (indicative), "yaktub ā" (subjunctive, jussive).

The infirm verb undergoes a number of modifications when the mood is subjunctive or jussive. The following is a brief summary:⁽²⁵⁾

i. Imperfect infirm defective verbs: This refers to imperfect verbs ending with the vowels "ī" or "ū" (al-manqūṣ), or with "ā" (al-maqṣūr). The forms in the various moods are as follows:

- indicative: no marker, e.g. yarmī, yad<ū, yardā.

- subjunctive: fatha (the vowel "a") is added as a marker to the final "ī" and "ū", making them "iya" and "uwa" respectively; no marker for verbs ending with a, e.g. yarmiya, yad<uwa, yarda.

ii. Imperfect infirm hollow verbs: This refers to verbs the trilateral root of which has "w" or "y" in the middle, i.e. as the second radical. This second radical is realised as "ā" in the perfect, but in the imperfect it is realised as "ū", "ī" or "ā", e.g. yaqūlu (root: q-w-l, perfect: qāla), yamīlu (root: m-y-l, perfect: māla), yanālu (root: n-y-l, perfect: nāla).

- indicative and subjunctive: The same final vowel markers as the strong verb, i.e. "u" and "a" respectively.

- jussive: Replacement of the medial vowels "ū", "ī", and "ā" with the corresponding short vowels: "u", "i" and "a" respectively, e.g. yaqul, yamil, yanal.

This brief outline indicates the intricacy of the paradigms that result from placing a particular verb in the various moods, an indication of the interaction of syntax and morphology at this level of linguistic realisation.

d. Voice: This category intersects with tense, moods, person and number, creating a complex of paradigms. This intersection is, in English, realised by the various elements that make up the experiential structure of the verbal group. In this respect, it is the Finite and auxiliary elements that realise number, person as

well as tense, leaving the Event element in the past participle form. In Arabic, all these categories operate on the form of the Event, and therefore a one-word form represents the fusion of various categorial modifications.

e. Number: This is pre-eminently a noun category, which is grammaticalised both in English and Arabic, and is reflected immediately on word forms. English makes distinction between singular and plural, while Arabic numerical status is tripartite: singularity, duality (alluding to two individuals) and plurality, which implies more-than-duality. Number, in both English and Arabic, is overtly marked and is transmitted to accompanying determiners and verbs (in Arabic it is extended to adjectives both in attributive or predicative functions). Moreover, number in Arabic interacts with gender to produce a subtle categorisation which requires, especially in the plural, specific markers, the description of which lies outside the scope of this study and can be found elsewhere (cf. for instance, Haywood and Nahmad 1965, Hasan 1980, Wright 1962 and Beeston 1970). Verbs in the imperfect can, as shown earlier, indicate number by the initial marker of the verb, e.g. "aktubu" (I/write) versus "naktubu" (We/write). A similar distinction is made in English present/simple verbs, e.g. "he writes" as opposed to "they write".

f. Gender: If we, as is the practice of some linguists (cf. Palmer 1971) disassociate gender and sex, thus rejecting the correlation between gender as a grammatical category and sex as a typology of entities in the real world, we can agree that English is a genderless language. In Arabic, gender is grammaticalised in

that there is a high degree of arbitrary assignments in which entities are either masculine or feminine. Although there exists only two genders in Arabic (Haywood and Nahmad 1965, Wright 1962 and Beeston 1970 mention a third, a common gender, i.e. the potentiality of treating the word as masculine or feminine at the user's (speaker's/writer's) discretion.

Changes in word configuration to suit gender requirement affect nouns (including adjectives), pronouns, deictic words (including determiners and numeratives) and verbs. Concord in gender between the head of the noun groups and its attributive or predicative adjectives, deictic words, pronouns as well as verb (perfect or imperfect) is essential. Concord between the head noun and its numerative deictic follows particular patterns whereby concord or "reverse concord" are applicable.

A common marker of feminine gender in Arabic is final "at" realised in orthography as "ta' marbūṭa". This marks a big number of singular nouns (including adjectives) e.g. sā'atun kabīratun [a big watch/clock], and can also be associated with nouns indicating female, i.e. the notion of sex (e.g. "mu'allim" [man-teacher] and "mu'allimat" [woman-teacher]. Other markers of feminine gender involves patterns of vocalic change, as in "abyaḍ" [white-mas.] versus "bayḍā'" [white-fem.], 'awwal [first-mas.] versus "ūlā" [first-fem.].

Gender markers affecting imperfect verbs in Arabic are realised by initial "ya-" and "ta-", indicating masculine (third person) and feminine (third person) respectively. The perfect is marked by a

final feminine "t", as in "kataba" [wrote, mas] and "katabat" [wrote, fem].

g. Case: Modifications due to case are restricted in English to personal pronouns. In Arabic, however, the influence is, morphologically speaking, diverse as well as extensive. Case contracts to nouns (including adjectives) syntactic markers, the full range of which distinguishes three kinds of case:

i. Nominative: This corresponds to Beeston's "independent status" and refers to nouns belonging to the central nominal constituents of the sentence, i.e. the subject occurring as the theme (mubtada') of the clause (provided it is not preceded by a group of verb-like particles labelled by Arabic grammarians as "inna" and its sisters), the subject of a verbal-clause (where the verb is in the thematic position) and the noun predicate (xabar) in a non-verbal predicate structure (provided the theme is not introduced by the special verbs labelled "kāna" and its sisters).

ii. Accusative: This corresponds to Beeston's "subordinate status" and applies to nouns functioning as theme/subject and noun predicate in the two exceptions mentioned in (i) above. It also applies to amplifications of the predicate including the object and some time adverbs.

iii. Genitive: This corresponds to Beeston's "dependent status" and applies to nouns placed after prepositions, or nouns in the amplifying position (i.e. the post-head "mudāf ilay-hi") in the annexation structure (i.e. "idāfa").

The markers realising these cases are complex. They take the form of variable vowel terminations (i.e. -u, -a, -i) and "tanwīn" (i.e. -un, -an, -in) added after the stem as well as particular vocalic change in the markers of the dual, the masculine sound plural and nouns that end with the vowel "ī".⁽²⁶⁾ If the noun is inflectible, then, according to Beeston's categorisation, it falls into one of seven types, each selects the syntactic markers differently to represent any of the three cases (cf. Beeston 1970, p.52 for a summary). Adjectives mirror the nouns which they amplify by taking their same case. Thus a noun in the accusative case will have its accompanying (attributive) adjective also in the accusative. Case affects Arabic pronouns, e.g. the connected pronoun "tu" [I] is the subject of a verb in the perfect (e.g. "qābal tu" [I met]), which becomes "ī" in the accusative (as an object, always preceded by "n", called "nūn al-wiqāya" [protection n] by Arab grammarians, e.g. qābala nī [he/met me]) and in the genitive (e.g. "kitāb ī" [my book]).

2. Modification due to juxtaposition

Certain words in Arabic undergo certain modification in form if they lie in close proximity (immediately preceding or following) certain other words. This modification affects some pronouns, nouns, verbs and prepositions.

a. Pronouns

i. The "connected" pronoun "hu" (third person singular accusative or genitive case) is rendered "hi" if preceded by a noun in the genitive (e.g. "hāḡā kitābu-hu" vs. "fī kitābi-hi").

Similarly the pronouns "hum" and "humā" are rendered "him" and "himā" for the same reasons.

ii. The "connected" pronouns "'alif al-muṭannā" (-a) [dual pronoun], "waw al-ḡamā'a" (-u) [masculine plural pronoun] and "ya' al-mu'annat" (i) [feminine second person nominative pronoun] are rendered "āni", "ūna" and "īna" respectively if the verb preceding is imperfect indicative.⁽²⁸⁾

b. Nouns: The modification here is purely orthographic. Nouns ending with "ta' marbuta" or "'alif maqṣūra" will have these letters changed into ordinary "ta'" and "'alif" respectively if the nouns enter in "idāfa" structure where these nouns are the head "muḍāf" (first term in the annexation) and a connected pronoun is the post-head "muḍāf ilay-hi" (second term in the annexation)⁽²⁷⁾

c. Verbs: This involves the Arabic infirm verbs. One modification is orthographic, whereby "'alif maqṣūra" is changed into an ordinary "'alif" if the verb is followed by relevant "connected" object pronouns. The rest of the modifications are morphological and affect perfect infirm and solid verbs when followed by the subject "connected" pronouns "tu", "ta", "nā", "ti", "tumā", "tum", "tunna". The modifications that result are:

i. Perfect infirm defective verbs where the third radical (the third element of the verb root) is "y" (ya') or "w" (waw) have this radical changed to "ay" and "aw" respectively, e.g. "ramā" [threw] (the root of which is r-m-y) becomes "ramay-tu" [I threw], "da'a" [called, invited] (the root of which is d-<-w) becomes "da'aw-tu" [I called, I invited].

ii. Perfect infirm hollow verbs (the second radical of which is either "y" or "w") have their second radical changed into a short vowel, "i" or "u" respectively, e.g. qāla [said] (the root of which is q-w-l) becomes "qul-tu [I said], "māla" [leaned] (the root of which is m-y-l) becomes "mil-tu" [I leaned].

iii. Solid verbs (verbs in which the second and third radicals are identical) usually undergo special change whereby a short vowel is inserted between the second and third consonants, thus cancelling the gemination of the two radicals, e.g. "massa" [touched], "masas-tu".

d. Particles: Particles that end with 'alif maqsura have this letter changed to "ay" when followed by a connected genitive pronoun, e.g. "<alā" [on] becomes "<alay-hi" [on him/it], "ilā" becomes "ilay-hā" [to her/it]. In addition the particle "li" [to, of] can sometimes be changed into "la" in these positions:

i. if followed by a connected (genitive) pronoun except "i" [me].

ii. sometimes before a non-verbal predicator when the preceding subject is introduced by the emphatic particle "inna".

iii. when introducing a main clause after conditional clauses having such subordinators as "law" [if-hypothetical] and "lawlā" [except for/that, had it not been for].

All these modifications to word forms, far more in Arabic than English, tend to increase the size of V as opposed to N if each form is to be taken as a separate word type. Lemmatisation can be a

useful procedure for unifying these various modified forms into one; however lemmatisation has to be approached with caution if a complete statistical account is to be formulated for the lexical usage in the entire corpus. Before we consider this any further, we would like to examine the second category of exceptions of the main rule that equates one word with one orthographic form.

B. Several words = 1 orthographic form

This refers in particular to the question of homonymy and polysemy. Radically different meanings may correspond to one phonic (or graphemic) reality, and the two terms (i.e. homonymy and polysemy) are used to describe two different perspectives of this phenomenon. In the context of our discussion, one might argue, on the one hand, that, for instance, there is one word "cross" having various meanings. On the other hand, one might equally forcibly argue that these meanings represent different words "cross". There are instances that are more complex than this one, but even in a straightforward instance such as "cross", the question is not as uncontroversial as it would seem. The two perspectives we have referred to can be summarised as follows:

1. Authentic homonymy supposes that between different meanings of the same expression there is neither a common core, nor an even continuity, as in "date" (= a small brown sweet fruit), and "date" (= time shown by, for example, the number of day, month and/or year). This renders difficult attempts to explain any one by the other(s) or to derive them all from one basic signification. Etymology, while useful in diachronic and comparative linguistic

studies and sometimes capable of justifying the relatedness of meaning (Lyons 1981, p.46), is of no value for our purpose. The true criterion, we believe, is the absence or presence of a semantic bond. For instance, when a form has two meanings and its use in sense a and its use in sense b correspond to two choices that are as distinct as if two different expressions are used, it then represents two authentic homonyms. Kempson 1977 uses the term "lexical items" to refer to homonymous forms and restricts the use of the term "word" to the phonological complex. To her, the paradigm run, runs, running and ran are different forms of one lexical item (and therefore different words). She characterises the form runs in the two sentences: "he runs the motorshow" and "he runs for Hampshire" as two separate lexical items and suggests having each in a separate entry in the dictionary.

2. A word with a particular meaning (a "lexical item" in Kempson's formalisation) does not always manifest a consistent meaning common to all occurrences of that item. An instance for this is the change of meaning in the lexical item "hold" in these two contexts "She held the baby" and "She held a half share in the business". For such an item one has to allow for vacillation of meaning from context to context, a phenomenon traditionally distinguished from homonymy by the term polysemy. The contextual determination and relatively general laws may allow passage from one meaning to another and allow us to foresee the variation.

What concerns us here is that there seem to be two alternatives: one is to say that the meaning of a certain lexical item is not constant - but context-determined and so varies from one context to

another. This implies avoiding setting up variations in type for a particular token and, as a result, will keep V at a minimum. The other alternative is to claim that the meaning a lexical item has is constant, but that sentences such as the two just mentioned contain two different lexical items: "hold 1" and "hold 2". The adoption of this approach will increase the size of V and, accordingly, the lexicon in which each item is characterised will be much larger than might have been anticipated.

The problem of homonymy and polysemy in word count has no general solution. It is possible, of course, to take one dictionary as the supreme arbiter; or, alternatively, a list of homonymous forms can be assembled and treated as separate words. Both solutions suffer from the disadvantage that in automatic production of word counts, sophisticated software is required to differentiate semantically the various meanings of homonymous or polysemous forms. Such software is still experimental and unless semantic networks are adequately developed, a considerable size of manual work has to be incorporated. Even when such manual work is attempted homonymous lists that are to be assembled have to be kept within reasonable limits, otherwise one would simply have to recognise as many different members of V as there are different meanings that homonymous and polysemous items can manifest. This will, in effect, render any attempt at word counting illusory and even useless.

It should be remembered that the main criterion that governs our attitude in this matter is the nature of the purpose of word count. In this respect Herdan (1966) believes that "whereas a count for

teaching purposes will favour a considerable degree of differentiation of forms according to meaning, a word count for purposes of establishing certain coding properties of language can afford to restrict such differentiation to a minimum, if not neglect it all together". (p.167).

App. 1.3.2 Relevant Criteria

In word count studies there exists a number of norms for establishing the size of V, none of which is perfect. This is because linguists' views may differ in relation to certain facts. Besides, investigations into the quantification of language can go into different directions. In the face of the diversity of methods available, we must insist in this project on this: that for achieving the aims of our quantitative analysis, efforts should be expended to make the norm explicit. For words to be members of V, formal criteria, more syntactic than semantic, have been employed. The following are adhered to in this project:

1. Form vs. meaning

Consideration of the dichotomy of form and meaning has been one of the concerns of modern linguists and we feel it is not appropriate to retackle the topic in detail within this thesis. It suffices to indicate that within the general word count of this project, form has been opted for as indicative of membership of V. Differentiation of homonymy and polysemy has been overlooked, except where it is an essential part of the analysis of connectives. In this case tagging procedures are performed when the focal task is establishing a semantic categorisation of connectives according to

function and textual range (see Ch.6 above).

By "form" it is here referred to the graphemic and grammatical configuration of the members of N. Accordingly, items that have identical configuration are listed under one entry.

2. Grammatical categories

Certain grammatical categories have been retained to further distinguish forms that otherwise get classified under one entry. These are the categories of person, tense, voice, number and gender. Two forms sharing the same base will be treated as two distinct members of V (two different types) if the distinguishing factor is related to any of the categories above. However, two categories have been overlooked and are here treated as non-influential in determining the membership of V. These are the categories of mood and case. This is justified on the basis that the intricacy of the mood inflectional system of the Arabic verb and the various types of case marking that affect inflectible nouns (including adjectives) will result, if these two categories are allowed to determine membership of V, in an unwarranted increase in the number of types as opposed to tokens. We have therefore been persuaded to consider these categories as irrelevant.

This decision, we realise, dictates the use of tagging procedures to help equate mood forms and case types together. Accordingly, a verb in the indicative, e.g. "yaktubu", will be treated as the same type as "yaktuba" (subjunctive) or "yaktub" (jussive). Similarly, infirm verbs such as "yarmi" (indicative) will be treated as the same type as "yarmiya" (subjunctive) and "yarmi" (jussive).

Nouns differentiated on the basis of their grammatical case within the clause, such as "kitābu" (nominative), will be included in the same entry as "kitāba" (accusative) and "kitābi" (genitive). Sound masculine plurals such as "mu'allimūn" (nominative) is equated with "mu'allimīn" (accusative and genitive).

An exception, as far as the equation of various case marking is concerned, is made to personal pronouns. Words such as "I" and "me" (English), "'anā", "tu" and "ī" (Arabic), are treated as five distinct types. It has been felt that equating these pronominals will not introduce a statistically representative image of the behaviour of these words. However the Arabic pronominal form "ī" as in "kitab ī" is equated with "nī" as in "qābala nī". Both forms are treated as one, since they are in fact the same pronoun if the "n" (nūn al-wiqāya [protection "n"]) is ignored. This is an old recognised grammatical fact of the Arabic pronominal system.

3. The influence of juxtaposition

It has been exhibited that certain words can have more than one form when placed in proximity with certain others. We have resolved not to consider juxtaposition as a determining factor of types as opposed to tokens. Accordingly, a verb such as "masas" [touched] (when occurring immediately before the "connected" pronoun "tu" [I]) is treated as a token of the type "massa". Similarly, the proposition "<alay" [on] in "<alay hi" [on him/it] is a token of the type "<alā".

These three criteria for selection of types are more operational in Arabic than English owing to the complexity of the syntactic and

morphological patterning of Arabic. It is felt, within the scope and possibilities of this work, that without setting these three criteria for the delimitation of types, the size of V will eventually grow so unmanageable that a) word count will fail to approximate reality, and b) contrastive statistical analysis with English will be inefficient and the results it may yield will be both linguistically and statistically implausible.

App. 1.4 Conclusion

Word count mathematics denotes, among other things, the determination, by counting, of the occurrences of words used in a linguistic text, or samples from it. This immediately brings into question the identity of the word as a unit of running text (or a unit of N). The delimitation of this identity is a necessary prerequisite, particularly in Arabic, since connectives such as "wa" and "fa", are traditionally written as parts of the following words. Any calculation of text length in terms of words would therefore require a working solution.

Furthermore, since statistical condensation of the tokens (members of N) requires the grouping of words under headwords (types), another identity has to be delineated. This is the status of word as a lexical unit, or a member of V.

In defining the identity of members of N (tokens), we have used four criteria: semantic, grammatical, criterion of internal stability and criterion of lexicon. Those criteria have been applied to Arabic and English and the word is defined accordingly.

After membership of V in a number of studies has been outlined, criteria have been set that are limited by pragmatic considerations and inspired by the fundamental requirement of investigating and contrasting two widely different and genetically unrelated languages (i.e. English and Arabic). One of the main criteria treats form, defined as the total graphemic and grammatical configuration of the members of N, as a determining factor in specifying types. The next criterion, directly related to the first one, is that of grammatical category. Forms of a certain base that are differentiated on the basis of one or more of the following categories are taken as distinct members of V: a) person, b) tense, c) voice, d) number and e) gender. Two grammatical categories have, however, been excluded: mood and case, which are particularly operative in Arabic. Accordingly, forms that are distinct morphologically on the basis of their mood or case inflections are to be equated and treated as tokens of one general type.

The third criterion that is adopted is the neutralisation of the morphological and orthographic changes imposed on forms by juxtaposition of certain words to others. Hence, Arabic words that are usually modified in form when placed in proximity to other words are equated as tokens of one general type.

The limitation of the last two criteria raises the problem of form equation during the processing stage of the corpora. For an analysis of this problem and the use of convenient solutions, refer to Chapter 5, Volume 1 of this thesis.

Footnotes to Appendix (1)

- (1) For an elaborate treatment of the semantic content of the word see Komlev (1976). Differences in the identity of the word across languages have been enunciated by Vladimír Skalička, quoted by Kramský (1969). According to Skalička, such terms as "word", "sentence", "case", "tense", "mood" are commonly used for different languages, despite our awareness of the fact that they are used in a rather different sense. The term "word", he maintains, means something else in Czech, German, French, and adds that we are content with an approximate identity. Skalička's pessimistic attitude is rejected by Kramský (1969, p.9).
- (2) For a comprehensive analytic bibliography of works on the linguistic concept of the "word" see Juilland and Roceric (1972).
- (3) Quoted by Kramský (1969, p.9).
- (4) Quoted by Kramský (1969, p.17).
- (5) Cf. review by Fowkes (1959). Bazell concedes that these criteria are not applicable to all languages to the same extent, and as examples he quotes Czech and Hungarian.
- (6) Gelb (1952), quoted by Barr (1976), maintains that the old semitic consonant scripts are in fact syllabic in character. This opinion seems to have had some influence on some books in general linguistics, e.g. Robins (1964/1980, p.97), Abercrombie (1967, pp.38 and 168, fn.6). Barr regards Gelb's arguments as "unconvincing or unsatisfactory" (p.74).
- (7) In normal writing these diacritics are dropped, leaving the consonantal skeleton by itself.
- (8) In this respect Crystal 1971 states "We cannot really define words by reference to this pause feature, as this is something which is only characteristic of abnormal speech. We surely have to be able to identify words under normal conditions, if the concept is to be really usable" (p.189).
- (9) Chomsky and Halle (1968, p.368) report that a similar situation to this is the common error that semi-literate writers of Russian make when they omit the space after a preposition.
- (10) Note that in certain Romanised versions of Arabic that investigators use, a hyphen is introduced to separate these words: katab-tu and kitāb-ī. (See for instance Koch 1981).
- (11) The default of OCP (Oxford Concordance Program) does this. The comma and the full stops are treated as space, and therefore can divide a string if any occurs within it.

- (12) This is the usual practice employed in the articles of the Daily Telegraph for instance.
- (13) The distinction between "synsemantics" and "autosemantics", as reported by Kramsky (1969), comes from A. Marty (1925). "Unter 'Wort' aber verstehe ich jedes Sprachmittel, das als besonderes Glied des Organismus der Rede empfunden und als besondere semantische Einheit behandelt wird. Ich sage nicht: eine besondere semantische Einheit ist. Denn wie schon früher bemerkt wurde, kann man dies selbst von den logisch begründeten Synsemantika nur in einem weiteren Sinne, im strengen Verstande aber bloss von den Autosemantika sagen."
- [By "word", however, I understand any linguistic means which is felt to be a special member of the organisation of speech and is treated as a special semantic unit. I do not say: which is a special semantic unit. For as was remarked previously, this can be said about logically based synsemantics only in a further sense; in the strict understanding however it can only be said about autosemantics].
- (14) Reported in Kramsky (1969, p.20).
- (15) Quoted by Kramsky (1969, p.23).
- (16) See a summary of these tests in Swaminathan (1973).
- (17) This brings out the vital distinction between a word and a sentence, and can add a further dimension to the difficulty of defining a "word".
- (18) A quotation and a short discussion is in Kramsky (1969).
- (19) Such a consideration is completely overlooked by Ager (1974 and 1977). Ager however resorts to the orthographic word in compiling his word frequency indices.
- (20) For a brief treatment of these approaches see Ferris (1983).
- (21) The word "independent" should not be confused with "orthographically autonomous". It here refers to the internal stability among its components.
- (22) Connectives that are identified and tagged in the Brown corpus (Francis and Kučera 1982, p.13) are coordinating conjunctions (tagged CC) and subordinators (tagged CS). No attempt is made to identify or tag intersentential connectives.
- (23) Mood in English, as it seems to me, is a semantic feature that is marked morphologically though this marking is receding and is perhaps restricted to American English or some old legalistic text (cf. Leech 1971, p.106). Mood of the Arabic verb is, however, a syntactic/morphological feature.
- (24) The energetic usually occurs in classical or elevated

language. It is formed by adding the termination -nna or -n called by Arab grammarians "al-nūn al-mu'akkada" [corroborative n] to the imperfect, as in "yaktubunna" [he (definitely) writes] (Wright Vol.1, p.61). This form, owing to its archaic use, is dropped from consideration here.

- (25) For a more detailed account, refer to Wright (1962, Vol.1) and Hasan 1980 (Vols. 1-4).
- (26) The marking system for those can be summarised as follows: 1) dual: a) nominative: the use of (-āni) at the end e.g. "kitābāni" [two books-nom.] b) accusative and genitive: the use of final "-ayni", e.g. "kitābayni" [two books - accus. & gen.]. 2) masculine sound plurals: a) nominative: use of final "-ūna" as in "mu'allimūna" [teachers-nom.]. b) accusative and genitive: use of "-īna" e.g. "mu'allimīna: [teachers - accus. & gen.]. 3) nouns ending with ī (called "manqūṣ" by Arab grammarians): a) nominative: no marking, e.g. nādī [club], b) accusative: "ī" is changed into "iya" or "iyan" when definite or indefinite respectively, e.g. "al nādiya", nādiyan". c) genitive: no marking if definite "al nādī", or omission of ī and replacement with "in" when indefinite "nādin". 3) a group of five nouns, often called the five nouns by Arab grammarians; those are "'abū, 'axū, ḥamū, fū, dū" used as head nouns in annexation structure [iḍāfa] (as in "abū ka" [your father], "'axū hu" [his brother], "ḥamū ha" [her father-in-law], "fū ka" [your mouth], "dū mālin [possessor of wealth, rich]. a) nominative: "ū" is retained in each noun, no further marking; b) accusative: "ū" is changed to "ā", as in "'abā ka"; c) genitive "ū" is changed to "ī", as in "'abī ka".
- (27) Note that the transliteration scheme used for coding the Arabic corpus used the symbol "T" for "ta' marbūṭa" and "t" for ordinary ta'. The symbol "V" is used to represent "'alif maqṣūra" while A represents "'alif mamdūda [lengthened 'alif].

Appendix 2

Approaches to Text-Based Studies

App. 2.0 Perspective

The content of this Appendix is a critical review of some of the major text-based approaches. The review has been intentionally compacted, but it is, we hope, sufficient to serve the purpose of sketching a profile of the available trends. A more detailed exposition could overload the arguments we are trying to introduce and will certainly go beyond the scope of the study. The trends to be reviewed include classical rhetoric, stylistic studies, semiotics, content analysis, "new rhetoric", discourse analysis and text linguistics.

App. 2.1 Introduction

Current trends that are favourable to the development of text-based studies and the development of text theories, though characterised, since their emergence more than a decade and a half ago, with profuseness as well as divergence, are not in actuality a new phenomenon. A chronological examination of the history of textual studies (attempted by Dressler 1972, Hartmann 1980 and Hatim 1981) allows us to generalise on the type of setting and factors that tend to promote the view of language as connected sequences of discourse. Such an examination identifies a number of approaches that, despite their fuzzy boundaries and their usual interrelatedness, can still be categorised in different groupings.

The aim of this appendix is to outline these approaches and

briefly examine some of their tools of investigation, their efficiency for analysis and their rigour of description in an attempt to characterise the approach that this project intends to utilise. Fundamental to this outline is the need to offer a comparison of approaches which can serve as a basis for subsequent presentation of the different and seemingly divergent paths of methodology that our project intends to follow in order to offer a clear and efficient description of the problem under investigation (i.e. the theoretical description of the textual role and the quantitative profile of connectives in English and Arabic and the practical application of the synthesis).

App. 2.2 Classical Rhetoric:

Any examination of text-based models should take into account classical rhetoric, which represents the first impulses towards laying down precepts for discourse. This preceptive movement started in ancient Greece, was transmitted to Rome, and thence to Medieval Europe.

By laying down specific directions, based on an analysis of current practice, rhetoric has enabled the experience of skilled public speakers to be conveyed to later generations in the form of direct suggestions for conduct (Murphy 1974). In general, the principles on which rhetoric operates can be summarised as follows (cf. Hartmann 1980):

- a) Rhetoric deals with language as real discourse;
- b) Its methodology aims at isolating categories and criteria

for effective communication;

c) It provides practical guidelines for one group of language user, the public speaker.

Greek rhetoric was not a single, homogeneous theory (Young et al. 1970 p.3). However, among the early contributions, Aristotle's "Rhetoric" stands out as the oldest extant textbook on the subject. Aristotle there defines rhetoric as the faculty of discovering all the available means of persuasion.⁽¹⁾

Roman rhetoric is traditionally associated with Cicero and Quintilian. Cicero defines rhetoric as a department of political science dealing with eloquence based on rules of art (Murphy 1974, p.9) Quintilian, a first-century Roman orator, wrote a textbook in which he outlined the rhetorical categories and conventions of the art of speaking. His distinction of five parts and three types of discourse remains valid today; it even foreshadowed much of later work on the specification of sociolinguistic variables of discourse.

The limitations of this approach (cf. Hartmann 1980, Hatim 1981) appear in the incapability of its static model to admit change and variety in time, context and subject matter. Additionally the principles of textual organisation are unsatisfactorily explored, and, most seriously, the correlation with situational factors are minimised. To these shortcomings is added the uncritical imposition of a universal pattern, derived from Classical Greek or Latin, on all other languages, no matter how differently each language can realise its rhetorical conventions. Indeed, recent research on intelligent behaviour has indicated that the human brain does not

generally conform to Aristotelian categories (cf. Krippendorff 1980 p. 9).

Related to the field of classical rhetoric is the field of exegesis. Its task is to interpret discourse with the aim of discovering the "true meaning" of passages. This approach has been criticised (cf. Hartmann 1980) on the basis that its methodology relies heavily, in its search for internal evidence, on close reading rather than on the linguistic analysis of relevant discourse factors. (2)

App. 2.3 Stylistics:

Stylistic studies have emerged from the need for a closer comparative description of different language varieties. Within these studies, the aim is to show what relationships are between the writer's intentions, the features that constitute his text and its reception by the reader. A number of schools can be recognised within this tradition, which Guiraud (1974) divides into four: descriptive (with representatives like Bally and Maurice Grammont), functional (Roman Jakobson and Ronald Barthes), "genetic" (Benedetto Croce and Leo Spitzer) and quantitative stylistics (Wilhelm Fucks, Yule, and Herdan).

One of the criticisms launched against stylistic studies is its subjectivity: views tend to be dogmatically formulated and the methodological applications can hardly be verified objectively. This is what led linguists (cf., for instance, Leech 1969, Fowler 1970, Crystal and Davy 1972, Enkvist 1973, Sinclair 1975, Widdowson

1975) to attempt to control and confine the intuitive nature of the speculations and judgements of literary critics. Other problems facing stylistics are summarised in Sinclair (1985) and include "the lack of principle in selecting a focus of description, the uneasy status of interpretations from stylistic evidence, and the difficulty of description of long texts" (p.17). Sinclair adds to those the problem of having "no theory in stylistics; the value of its observations is related strictly to the results of individual studies".

App. 2.4 Semiotics

Real progress in studying the intricacies of the communicative processes was achieved when the properties of signalling systems, both natural and artificial, started to be scientifically investigated within the field of semiotics. Among the early contributors to this field are the philosophers Charles Peirce, Charles Morris and Rudolph Carnap (Crystal 1980) as well as the linguists Saussure and Jacobson.

In its oldest sense, semiotics refers to the study within philosophy of sign and symbol systems in general. Since the verbal signs have always been in the limelight, reflection on signs was for a long time inseparable from reflection on language.

The linguistic approach to semiotics is the one outlined by Saussure and followed by Hjelmslev, the French structuralists and a host of other linguists. Saussure adopted the term semiology to refer, in his own words, to a "science that studies the life of signs within society". Semiology would show what constitutes signs,

what laws govern them" (Saussure 1960, p.16, his emphasis). Since language is the most important sign system and nothing besides language is more appropriate to make nature of the semiological problem understood (loc. cit.), linguistic semiotics starts with linguistics in order to establish, by analogy, a general model of semiotics (cf. Koch 1971 p. 24 quoted in Nöth 1978 p.22). Such a model can then be applied to other sign systems. Barthes (1967), recognising the heuristic priority of linguistic analysis in semiological research, concludes that non-linguistic semiotics should be regarded "a second-order language" (p. 11) which does not exist independently of language.

Morris, perhaps among the most influential contributors to semiotics, introduced in the 1930s a number of suggestions that have enjoyed popularity in linguistics. Being a logician as well as philosopher, he first introduced logic within the framework of semiotics, and formulated a series of clear distinctions, such as those between designatum and denotatum. Similarly, he distinguished three dimensions of the sign: semantic, syntactic and pragmatic. According to him, the relation between signs and designata or denotata is semantic; that between signs themselves is syntactic; that between signs and their users, pragmatic.

After Morris, numerous competing conceptualisations have been proposed, in particular for the pragmatic nature of the communication process. What they have in common is a comprehensive picture of the constituent parts of the communication act. Within those models, Hartmann (1980 p. 14) recognises seven components:

speaker or sender, audience or receiver, reality or object/events, message or text, code or language system, channel or medium, context or situation. The characterisation of each of these components and their interrelationships have been studied closely and, within the last few years only, a substantial size of research has been devoted to their investigation, the survey of which is irrelevant here. It suffices to conclude that these semiotic models have been used as a hold-up of all the major linguistic and non-linguistic factors of the communicative process.

The semiotic approach, despite the existence of an extensive body of research (Eschbach 1974, quoted in Nöth 1978, includes in his bibliography of semiotics more than a thousand titles), remains more a project than an established science (cf. Ducrot and Todorov 1979 p. 90). This is because the picture that emerges from the variety of papers and monographs confirms that semiotics is a heterogeneous field. Nöth (1978) mentions semiotic approaches to animal communication, theology, epistemology, basic research in mathematics, film analysis, psychiatry, communication theory, architecture, aesthetics, mass communication, advertisements, and literature. The question that has been raised (cf. Eco 1972 pp. 28-44) is where semiotics actually ends and where other disciplines of research begin. Virtually, the whole spectrum of humanities and social sciences, including efforts to improve the political and social conditions of life, is concerned with symbols, code, meanings, messages, their functions and their effects.

In addition, there exists some uncertainty as to the basic principles of semiotics as a science. Peirce's original conception

states that non-linguistic signs can be employed for locating the place of language among them. However, precise identification of these signs can prove to be a difficult and impractical approach. On the other hand, language can be employed for studying other sign systems (Saussure's approach, cf. also Barthes 1967). The result could be an impractical imposition of a linguistic model on phenomena that differ from language; semiotic activity can thus be reduced to denomination (cf. Ducrot and Todorov 1979).

App. 2.5 Content Analysis

Empirical inquiries into communication content have their root in studies in theology in the late 1600s, when the Church was worried about the spread of non-religious matters through newspapers. However, as a term in linguistic studies, content analysis is no more than 50 years old. Modern pursuits in content analysis are markedly different from their ancient roots in aim and in method. Krippendorff (1980) sees three marks of distinction. First, the pursuit of content analysis is fundamentally empirical in orientation, exploratory, concerned with real phenomena, and predictive in intent. Second, content analysis has claimed to transcend conventional notions of content as an object of concern and is intricately linked to more recent conception of symbolic phenomena. Third is the claim of the proponents of content analysis of developing a methodology peculiar to this field, which enables researchers to plan, communicate, and critically evaluate a research design independently of its results.

Content analysis was given a powerful impetus through the work

of Harold Lasswell and his associates (see for instance Lasswell et al. 1952, Lasswell et al. 1965). Their contribution was devoted to basic issues of sampling, measurement problems, reliability and validity of content analysis, all within the framework of analysing propaganda. Lasswell made his theoretical focus the political myth, stating that a key symbol is a basic term of the political myth. He then sets out to investigate and analyse the facilitating or retarding factors associated with the appearance of symbols and myths in the political realm.

The nature of the tasks of content analysis is included in the definition proposed by Krippendorff (1969 p.11) "the use of replicable and valid methods for making specific inferences from text to other states or properties of its source". The term "text" is used here in its most general meaning to stand for pattern in any communicative mode, and the term "source" is also generalised to denote any system to which inferences may be made from "text". Krippendorff, himself a specialist in communication theory, suggests that content analysis should not be restricted to the analysis of verbal data. Other non-linguistic forms of communication, such as painting, facial expressions, music, can be subjected to the same kind of analysis, even though "content" is not always conventionally attributed to these forms. Among the other questions of communication that content analysis attempts to study are the nature of the message, properties of the source and receiver, transmission of the data and data making or coding.

Proposals for the study of the relations between linguistic

expressions and real world phenomena are offered in Hays (1969). Hays puts forward an outline of what a theory may have to look like that would link natural language utterances to properties of its source. The key to this approach lies in his notion that in any linguistic exchange communicators utilise a substantial amount of non-linguistic background knowledge. Informal conversations and formal official exchanges are similar in that messages are not only interpreted in terms of this background knowledge, but also add to it continuously.

The abundance of research in this area leads one to expect that content analysis has evolved into a scientific method that promises to yield inferences from essentially verbal, symbolic or communicative data. However, the picture that emerges is an inverse to this expectation. A number of weaknesses have been spotted concerning the methodology and theoretical background of content analysis. First, the quantitative apparatus that Lasswell and his associates, and later his followers, developed and implemented has been found too crude to allow systematic generalisations. Smythe (1954) calls it "the immaturity of science" in which objectivity is confused with quantification. Second, it has been pointed out, even by proponents of content analysis (cf. Krippendorff 1969), that works of a methodological or theoretical nature as well as papers that attempt to elaborate critical issues or offer proposals for direction are distressingly hard to find, despite a number of attempts to rectify this disparity in the literature (for instance the papers in Grebner et al. 1969 and in Stone et al. 1966). Third, the type of data that content analysis proposes to study includes a

wide range of non-linguistic forms of communication, thus rendering a unified theory, or a set of procedures for analysis, hard to envisage. Fourth, there is no generally accepted agreement on a positive notion of "content". Some of the traditional ways of conceptualising "content" have posed some analytical and methodological problems which have, in some cases, restricted the content analyst's task. It is for these reasons that an adequate and systematic analysis of text, particularly where such textual principles as cohesion and coherence are concerned, is hard to pursue within the boundaries of this field.

App. 2.6 "New" Rhetoric

The introduction of "New Rhetoric" as a text-based approach has been seen as a basis for a method of improving competence in writing. Many composition teachers, especially in American high schools and colleges, were dismayed by the intellectual emptiness and practical ineffectiveness of composition courses (Young et al. 1970). Added to that is the feeling that new methods of grammatical analysis and pattern practice, and sophisticated approaches to punctuation and usage came nowhere near providing the basis for a coherent and comprehensive method. The solution that was suggested is a return to rhetoric with the conviction that it was potentially an important part of a college student's education.

But rhetoric in the Aristotelian sense posed a number of problems (see App. 2.2 above). Flesch (1962/1976), for instance, put an emphasis on the claim that Aristotle's rhetorical principles are moulded for Greek, not for English. A new concept of rhetoric

is thus needed which can cater for the requirements of the situation and meet pedagogical necessity and sufficiency. One such conception is put forward by Young et al. (1970), who defined rhetoric much more broadly than it had been defined for years. According to them, rhetoric is concerned primarily "with a creative process that includes all the choices a writer makes from his earliest tentative explorations of a problem in what has been called the 'prewriting' stage of the writing process, through choices in arrangement and strategy for a particular audience, to the final editing of the final draft" (ibid p. xii).

Such a conception is first concerned with prewriting or "invention", i.e. discovering information, forming concepts, seeing relationships, and analysing and solving problems prior to the act of communicating. Additionally, a special interest is expressed in the relationship of writer and reader (see, for instance, Flesch 1969/1976 pp. 26-36, Corder 1979 pp. 75-90, Eisenberg 1982 pp. 5-6, Huckin and Olsen 1983 pp. 47-58, Jordan 1984 p. 3). This relationship is divided by Corder (1979) into referential relationships, active relationships and personal relationships. The writer's task as seen by Young et al. (1970) is to discover his reader's values and interests and what his readers know about his topic, to understand and appreciate the prerequisites for his interaction with his reader's, in order to be able to accomplish his goal of informing the reader, strengthening his convictions, or changing his mind.

Emphasis is also laid on the linguistic choices the writer can make both during the process of composing his draft or revising it.

Within this concern, problems of structure, coherence, focus and loading are discussed. Another emphasis is put on the appropriateness of style. An intelligent style is regarded as an efficient and effective way of solving the sequence of problems that the process of writing presents. Young et al. (1970 p.361) reject the two popular notions: a) that practice alone is sufficient to develop an intelligent style, and b) that natural talent for writing is essential. They believe, rather, that an intelligent style can be taught provided that the necessary tools for achieving this operation are available.

Another concern that the proponents of the new rhetorical methods have expressed hinges on the effectiveness of scientific and technical communication and of promoting the communication skills of technical professionals (see, for instance, Eisenberg 1982 and Huckin and Olsen 1983).

This growth within new rhetoric has, despite its size and its applicational value, been characterised, so far, with a lack of genuine appraisal of the techniques used. We are still waiting for the development of a mechanism that offer an integrated evaluation of the effectiveness of the approach in terms of desired results, time, cost, etc. Until such a mechanism is integratively developed and scientifically tested for validity and reliability, more work is needed in both theory and application.

App. 2.7 Text Linguistics and Discourse Analysis

The two most prominent approaches to text analysis are text linguistics and discourse analysis. Both call for linguistic

investigation into beyond-the-sentence grammar, which identifies and describes supra-sentential linguistic structure in written and spoken texts, and both are strongly motivated by a desire for understanding how meaning is attached to utterances and how "sentences" combine into larger units to form texts. This analytic concern does not exclude consideration of the structure of individual sentences in order to illuminate how textual constraints affect the choice of individual lexical, exophoric or anaphoric items, inter-clausal linking devices, deletion possibilities, etc., within a sentence.

Despite the difference between the two approaches in aim and methodology, they are, within the context of this review section, best considered together. Both share the assumptions a) that it is the text where linguistic and extralinguistic categories correlate, b) that investigations that aim to characterise the textual patterns must operate beyond the limits of the clause and sentence, c) that the relevance of the text to its context is of paramount importance, and, d) that textuality (see Sec. 3.2 in Ch.3 above for a definition) is a more realistic notion for capturing communicative events than the more narrowly conceived notions of phonemicity, grammaticality and semanticity.

The differences referred to between discourse analysis and text linguistics are reflected in their aims and methodology. According to Hartmann (1980 pp. 17-18) discourse analysis starts with the outer frame of the situational context and works inwards to find out which verbal features correlate with the specific communication

setting. This has a history in the ethnographic approach of American methodology and British sociology. On the other hand text linguistics has as its main concern the development of text grammar, which starts from within the linguistic patterns of the message and describes how they might be used in certain contexts. This approach has its roots largely in the European deductive linguistics. One other difference of task is the interest of discourse analysis in behavioural interaction, whereas text grammar is more concerned with such manifestations as textuality: its nature, and various facets and parameters.

The two approaches have to be considered complementary to each other. Both view from two different perspectives "the single, but previously not explicitly articulated fact that messages are encoded as (or translatable into) discourse, discourse is realised as text, and text must be organised into a pattern to be decodable" (Hartmann 1980 p.18). Since the scope and aims of our study lies within the realm of text linguistics, only a sketchy picture will be offered below for discourse analysis. A more detailed account of text linguistics is offered in Chapter 2.

The global study of discourse has led to the establishment of a number of models, each examining a certain dimension (for a historical review see Coulthard 1976, Werth 1976, Coulthard 1978, van Dijk 1985). One fundamental issue that the models attempt to account for is the interrelation between linguistic form, semantic interpretation and pragmatic use in order to understand the interactional behaviour patterns of people communicating. One immediate concern of the models, therefore, is to probe the

speaker's/writer's intention in order, for instance, to distinguish and to identify and describe discourse functions and offer an explanation of how they are realised and interpreted. The models, in other words, seeks to answer two fundamental questions: "How does the speaker encode his function in grammatical form, and how does the listener derive the correct function?" (Coulthard 1976, p.76). This covers, among many other aspects, the distinction between direct and indirect speech acts, or between literary meaning and irony, and the establishment of the type of felicity conditions and conversational maxims as well as the extent of the pressure they exercise on the interaction and turn-taking.

The diversity of the tasks that have been investigated is evident from the recent flood of projects and publications, which is a strong indication that researchers have found the new field worth exploring. Discourse analysis has been applied to the study of classroom interaction (Sinclair and Coulthard 1975), intonation (Brazil 1975), literature (Coulthard 1977, Burton 1978, Korpimies 1978), acquisition of language (Hatch 1978), conversational structure (Keenan 1974, 1975; Widdowson 1979, Burton 1980). Other studies investigated the close link between discourse analysis and teaching language as communication (Widdowson 1978) and more specifically to the teaching of language for specific purposes.

Footnotes to Appendix (2)

(1) Aristotle divides such means into artistic, or that furnished by the speaker, and nonartistic, or that furnished by external evidence. Three kinds of persuasion are then named: ethos, arising from the speaker's personal qualities, pathos, arising from the audience's emotions, and logical proof, depending upon argument. Types of speeches are divided into three: forensic (judicial), deliberative (political) and epideictic (occasional), depending upon ends, times and subjects. Aristotle discusses the relation of audience to the speaker in Book Two, beginning with the emotions that a speaker may exploit to achieve his purposes. Diction and arrangement are discussed in Book Three. It is argued that good style, whose essential quality is clarity, must be appropriate to both the speaker and his cause, and that metaphor is useful because it makes comparisons quickly. As for arrangement, Aristotle maintains that a speech consists of only two essential parts: statement and proof.

(2) However, refer to Jamal-al-din (1980) for an impressive account of the linguistic orientation of Arabic exegetic works.

Appendix 3

Criteria and Requirements for a Definition of Text

The content of this Appendix is related to Chapter 3 and is intended to outline a number of criteria and requirements for a delimitation of the entity "text". These have been set up in order to eliminate, or at least minimise, the weakness that has observed in some conceptions available in the literature (see Chapter 3 for a short review).

The criteria that we propose are:

1. Simplicity

A too complex explication will burden the definition, blur the theoretical construct and reduce its viability for our research.

2. Consistency

An explication, as a theoretical construct, should maintain a recognisably sufficient degree of consistency. A construct that is employed in a variable fashion and therefore offers a different denotation each time it is utilised cannot provide an adequately clear image and will tend to confuse the issues it is used for and render them implausibly vague and highly questionable.

3. Generality

A delimitation of the construct should have the quality of being adequately general for purposes of the description of text organisation and text utilisation. A too specific definition will

suffer from limitation of coverage, thus leaving outside its boundaries a number of units that can otherwise be safely called "text".

4. Operationality

This is a decisive requirement, and one that should be carefully considered particularly when empirical research is attempted. A delimitation that proves to be operationally inefficient will minimise the chance of producing a proper treatment of the issues in hand.

5. Human Plausibility

For this we incorporate a number of factors that we can bring to the construct and that aim to specify a borderline. These include tolerance, prior knowledge of the participants (actual or potential), the situation and the typology of text in use. A sequence of sentences with an arrangement and organisation that defies human plausibility will disturb and can therefore halt, temporarily or permanently, the act of communication (see Chapter 3.2.2). We are not trying to replicate intuition here. Intuition is only a heuristic and does not constitute a valid primary datum (cf. Crystal 1971, Beaugrande and Dressler 1981, see also further discussion of "intuition" in Chapter 5).

Appendix 4

Manual Input: Transliteration of Arabic

App. 4.1. Preliminaries

It has conventionally been accepted as a norm by the majority of investigators and operators (in linguistic enquiry, computer work, bibliographic control systems) to use the Roman script as a standard to which all the repertory of graphic signs in different languages is to be subjected. Standardisation methods and codification procedures have been attempted to a variable degree of success, and was encouraged by the influence of modern mass media and the desire of nations to inexorably probe into the culture of each other.

The content of this Appendix reports on the requirement for Romanisation in this project, comments on some problematic features of the Arabic script when used as an input for a computerised analysis, and discuss the overall nature and scope of our Romanisation scheme. This section is regarded an integral and essential part of the project for two main reasons: a) it describes the set of possibilities that have been exploited for text processing purposes, and b) it proffers a specification of coding, which perspective users of the Arabic corpus may extend or capitalise on. In short, the account on Romanisation serves as a condensed "manual" to accompany the Arabic corpus. This account will start with some definitions and introductory comments related to script conversion in general and Romanisation in particular. Then we discuss some problematic aspects in Romanising Arabic and

the systematic and functional requirements in this project. Finally we discuss general features of the scheme developed here.

App. 4.2 Some Definitions

The "term" script has a number of denotations. In this brief account it refers to the set of conventional graphic signs having distinctive shapes and designed to give visual representation to the elements of a writing system (Wellisch 1978). This set may comprise some or all of several subsets of characters, each of which has a certain function. In general, some characters have a semantic function, e.g. numerals (representing numbers), Chinese logograms (representing functions) and such logograms as % and &. Other characters may have a non-semantic function, e.g. letters (representing phonemes), syllable signs (representing a group of phonemes), diacritics and punctuation marks.

Following Wellisch (1978), we would like to distinguish between three uses of script:

1) Orthography: the most common use of script is for the orthography of a language, where it is a) uniform, designated to apply to all possible words or sentences, b) integrated, i.e. it obeys rules of the writing system,⁽¹⁾ c) standardised, i.e. it is the product of convention or the general consensus of its users.

2) Transcription: This is a specialised use that refers to the representation, in the characters of a dominant script, of the phonemes of a language written in a dissimilar script (or not written at all).

3) Transliteration: In this use the graphemes of a source script are converted into graphemes of a target script without particular regard to pronunciation, and also, at least in the strictest sense, without either adding or deleting any graphemes that are not present in the source script. This latter restriction has, however, to allow for a certain degree of latitude in romanising Arabic script, a consideration that will be elucidated later. The need for a particular system of transliteration is prompted when the faithful reproduction of the graphemes of a source script is technically unfeasible, or when certain purposes are more conveniently served by a graphemic representation in the target language. Both these factors are prevalent in our project.

App. 4.3 Romanisation and the Arabic Script

Romanisation refers to a form of conversion in which letters of the Latin alphabet are made to represent a language script that uses other characters or signs. In the case of Romanising Arabic script, there has been proliferation of schemes introduced by linguists and Orientalists as well as librarians and bibliographers. Almost every Western Arabist since the Renaissance, and every Arab linguist writing for Western readers since the development of comparative linguistics, has found it necessary to devise his own scheme, or, at least, has tried to add, delete, or change a few characters of another's transliteration scheme so as to make it more efficient and appropriate for his purposes.

The variations in the schemes reflect certain incompatibility

between the Arabic and Latin alphabets. The main sources of incompatibility are summarised below (see also Appendix 1): (2)

1. Arabic has letters that represent sounds not available in Latin or in modern European languages, and are not even approximately expressed by any Roman letter or letter combination. There are two ways to represent such letters: a) a nearest letter combined with a diacritic such as a dot, an underscore or a bar, e.g. t and h to represent ṭ and ḥ respectively; b) by using phonetic characters borrowed from the IPA set e.g. x and q to represent ṭ and ḥ respectively; c) by using two Latin characters, e.g. kh and dh to represent ṭ and ḥ respectively; d) by using some non-standard characters that may exist in the Latin or IPA sets, e.g. ' and 9 to represent (hamza), and ʿ respectively. Preference on the part of linguists, librarians, bibliographers and other users for one type of representation than others and lack of co-ordination of efforts results in variations in transliteration schemes.

2. In Arabic some phoneme values are represented by more than one character. Most notable of these are the phoneme /a:/ represented by the 'alif, and the glottal stop represented by the hamza. The first is normally represented by l, but also by ʿ and ʾ (respectively 'alif, 'alif maqsura and dagger 'alif) and the second can have a number of shapes ء أ إ ؤ. Another phoneme, /t/, normally represented by the letter ṭ, may be represented at the end of certain words as ṭa (ta' marbūṭa). The use of these letters may either be determined by the position of that phoneme in relation to others in the word, or by syntactic factors such as case, mood or gender. Sometimes the choice may only be explainable on historical

basis (such as the use of dagger 'alif). In Romanisation, differentiation of 'alif and hamza are usually neglected. But representation of ta' marbūṭa presents some problems. Some schemes use an ordinary t to represent it, particularly when followed by a case marker, others prefer to replace it by an h (Beeston prefers a postscripted h) and neglect any case marker, while the majority prefer to ignore it.

3. The short vowels /a/ ,/u/ and /i/ and other morpho-phonemic features such as "tanwīn", gemination, pause, "wasl" are represented graphemically by diacritics that are usually absent in modern printed material (see Appendix 1 for more detail). They are usually mentally substituted by a person who knows the language and can construe the meaning of words from their context. The representation of these diacritics in transliteration schemes is not consistent.

4. The representation of the definite article (ال) has been problematic. Some schemes prefer to represent its full value as "al" or "Al". Others prefer to use its spoken form which, with the presence of wasl, can either be reduced to "l" or lost by assimilation with the first consonant in the next word. The choice depends on the nature of this consonant. In this position consonants have traditionally been divided into two groups: "sun letters", with which l of the definite article is assimilated into one geminated letter, and "moon letters", which do not allow assimilation. For instance:

kitābur-rajuli [The man's book] (Here "al" is assimilated in the spoken form with the next word producing a geminated rr)

kitābu-l-binti [The girl's book] (Here "al" is not assimilated, but appears in a weakly pronounced form)

5. The two diphthongs that exist in Arabic ay and aw are graphemically represented as a y or w, a diacritic, "fatḥa", being assumed on the previous letter. These are represented either as "ai", "au" or "ay", "aw" respectively.

6. Two characters are not represented in Romanisation. The first is " 'alif waw al-jamā'a", a "silent" 'alif used after the connected pronoun ū when it is in final position. The second is ﻻ where two characters ﻻ and ﻻ merge together to form one symbol. While the first is not represented at all, the second is represented by two characters lā.

Having specified the problematic features of Romanisation of Arabic, we now consider the requirement of a scheme that can be followed in inputting the Arabic corpus.

App. 4.4 Transliteration Schemes

App. 4.4.1 Requirements

When we look at what a transliteration system can achieve, we have to consider and distinguish between three types of requirements. These will be discussed under the headings of field of application, systemic requirements and functional requirements. (3)

App. 4.4.1.1 Requirements of Field of Application

Transliteration as a technical operation is necessary for two types of application. The first is the assembling of a corpus in a machine-readable form ready for computerised linguistic analysis. The second application is general in nature: to code Arabic terminology, illustrative exemplification, bibliographical information and other types of textual excerpts used in the presentation of this study when discussing connectives, cohesion and textuality. Requirements for these two applications do not necessarily conflict with each other; one transliteration scheme may be designed to suit both.

App. 4.4.1.2 Systemic Requirements

These are inherent in the structure and form of the transliteration system. Broadly they fall into four types: general applicability, uniqueness of representation, and simplicity.

a. General applicability: The transliteration system is designed and implemented in such a way that the computerised Arabic corpus can be used for further research by other investigators. The system therefore is not only envisaged for this particular project, but is intended to be of general applicability, or at least can be such that modification to suit further applications or to conform to researchers' particular needs does not constitute an obstacle.

b. Uniqueness of representation: Since the corpus is made up of written texts, one practical requirement of the computerised transliteration scheme is a one-to-one relationship between each

grapheme in the Latin alphabet and grapheme in the Arabic script.

c. Simplicity: The scheme should not be too elaborate for its application, nor too simple as to create ambiguity. On the other hand, the scheme must enable any user to read or write it with a minimum of effort. This can be achieved when the Arabic graphemes are converted following as closely as possible the conventions of the Latin script. Multiplicity of letters to represent certain Arabic graphemes is to be avoided, as it tends to cause unnecessary complication and ambiguity.⁽⁴⁾ In short, the scheme should combine comprehensiveness of representation with clarity and simplicity of use.

App. 4.4.1.3 Functional requirements

These requirements affect the mode of operation of the planned transliterated system. They include pronounceability and reversibility.

a. Pronounceability: The scheme should attempt to maintain as far as possible the demands of pronunciation. A transliterated word should be pronounceable by a user who has never seen it before. To achieve this the transcribed letters should, as far as possible, be pronounced as they are pronounced in the target language. In addition, short vowels and other diacritics have to be represented and supplied to the transliteration to make it pronounceable.

b. Reversibility: A transliterated system should ideally make it possible to reconstruct immediately and unambiguously in the Arabic script any word that has been converted to the Latin alphabet. This

is indeed the only method for full and unambiguous identification of the original form of transliterated words. It also facilitates to a considerable extent automatic reconstruction of Romanised Arabic in its original script by using appropriate software (or by such machines as the Lasercomp).⁽⁵⁾

Having specified the requirements of a transliterated scheme for this project, we now consider the features of the system we have devised.

App. 4.5 The Transliteration System in this Study

App. 4.5.1 General Features

In order to satisfy the requirements mentioned above and, at the same time, resolve the problems that the Arabic script brings to a transliterated system with a maximal efficiency in application in this and in future work. The following are some general features of this system.

1. The system comprises two schemes: conventional and computerised. The conventional scheme is intended to code all Arabic words, sentences or text excerpts that are dispersed in the body of this thesis. The computerised one is used for coding the Arabic corpus on the computer and in the various stages of processing the corpus. It is also the scheme in which all Arabic computer output is to appear (including microfiche appendices). Hence the first scheme is general in nature while the second is more specific in its application.

2. Both schemes are fully vocalised and are therefore fully

pronounceable. However, there are differences in the representation of those Arabic phonemes/graphemes not available in the target language. The conventional scheme uses more traditional symbols, and is therefore easy to read, which is the main purpose in its application. The computerised scheme uses its own symbols and is therefore less readily pronounceable. In both cases instructions are required for users who are not familiar with the Arabic graphemic/phonemic system.

3. Both schemes represent the Arabic graphemes on a one-to-one relationship. This satisfies the requirement of uniqueness of representation and reversibility. However, the computerised scheme is more comprehensive than the conventional one in reflecting the nature of the Arabic writing system (see below).

We now consider briefly each scheme discussing its peculiarities and commenting on its efficiency in relation to its envisaged application.

App. 4.5.2 The Conventional Scheme

The conventional scheme is a modification of Wehr's system used in his authoritative dictionary. The following are some distinctive features:

1. Consonants:

a) Only one shape for the hamza ('). It is transliterated wherever it occurs, even in initial position (where some schemes prefer to elide it). Accordingly, the madda in initial position is

represented only by the long vowel (ā) following a hamza: 'ā.

b) Arabic consonants that do not have a phonemic representation in the Latin alphabet are transliterated as in Table (App. 4.1) below.

2. Vowels:

a) Short vowels are included for ease of pronunciation. These are of two types: internal, constituting part of the lexical structure of the word, and external, indicating case or mood. Great care has been exercised for ensuring accuracy of representation.

b) Long vowels are represented as "ā", "ū" and "ī" while diphthongs are represented as "ay" and "aw".

3. Diacritic

a) Shadda is represented by doubling the consonant.

b) Tanwīn (one type of flexional endings) is represented as "un", "an" and "in" joined at the end of a noun without any necessary indication by way of postscript or subscript.

4. Other Features

a) The definite article is represented as (al).

b) No provision is made for representing "hamzat al-wasl" or weakening the definite article. Hence no distinction of sun/moon letters is necessary. This is a trade-off between pronounceability and simplicity of representation and is justified on the basis that the scheme is basically broad graphemic transliteration and is not

narrow phonemic transcription.

c) "Ta' marbūṭa" is treated in two ways. It is represented by (t) in sentences and text excerpts, where case marking dictates its use. However, where it appears in isolated words, we have dropped it.

App. 4.5.3 The Computerised Scheme

This is a comprehensive scheme with elaborate representation of the Arabic writing systems. Some of its main features are:

1. General

a) The scheme makes use of the ASCII character set. Since the number of lower case characters is not sufficient to represent all the variations in the Arabic script, use is made of upper case characters.

b) According to (a) above, no provision for capitalisation has been made. Arabic script does not have a system of capital letters.

c) Punctuation marks have been preserved. However, Arab writers in general, and this is reflected in the corpus texts in particular, are not consistent in their use of punctuation marks. A comma is sometimes used where a full-stop is expected and semi-colons are non-existent. Questions are occasionally not followed by question-marks and use of dots and full-stops is rarely systematic. In coding the corpus we have resorted to the following pragmatic decisions:

i. Full stops are preserved wherever they are used except where a

question mark is expected or where its use is erratic (a misprint).

ii. Commas are generally preserved except where its use is erratic or where a full-stop or semi-colon would be expected. Commas are also introduced where necessary, mainly where consistency in punctuation is required.

d) The system is fully vocalised. Great care has been taken to ensure accuracy. A number of authoritative dictionaries have been consulted in checking word internal vocalisation: Al-Qamūs Al-Muḥīṭ, Muxtār Al-Ṣaḥāḥ, Al-Munjid and Wehr's Arabic-English Dictionary. External vocalisation, indicating flexional marking, has been checked carefully. Where two readings are possible, the most common one is selected and consistently used.

2. Consonants

a) For the sake of simplicity, no diacritics have been used (dots or bars) as in the conventional scheme. Instead a separate character is used for each grapheme not available in the Latin alphabet. Table (App. 4.1) summarises the differences between the transliteration of consonants in the two schemes.

b) All forms of hamza have been preserved and represented as in Table (App. 4.2).

Grapheme	Conventional	Computerised
ث	t	C
ه	h	H
خ	x	x
د	d	v
س	s	S
ص	s	c
ض	d	D
ط	t	p
ظ	z	Z
ع	<	e
غ	g	g
ق	q	q

Table (App. 4.1) Schemes adopted in the project for transliterating consonants

Grapheme	Computerised
ا	Q
ؤ	O
ك	G
ي	J
أ	E
إ	Y

Table (App. 4.2) Shapes of the hamza

Grapheme	Name	Computerised
ا	'alif	A
ﺀ	dagger 'alif	'
آ	'alif maqsura	V

Table (App. 4.3) Shapes of 'alif

Diacritic	Name	Computerised
ﻻ	two dammas	W
ﻻ	two fathas	@
ﻻ	two fathas on alif	N
ﻻ	two kasras	M
ﻻ	shadda	# (placed after the geminated consonant)

Table (App. 4.4) Some diacritics

3. Vowels

- a) Short vowels are represented as u, a and i.
- b) The shapes of the alif (as the long vowel /a:/) are preserved. Silent alif (used after the third person masculine gender plural number connected pronoun u when in final position) is represented by the character (L) (see Table App. 4.3).
- c) The long vowels َ and ِ are represented as (U) and (I) respectively.
- d) The two Arabic diphthongs are transliterated (aw) and (ay).

4. Diacritics

A special character is reserved for nearly each diacritic as in Table (App. 4.4)

5. Other Features

- a) The definite article is represented as Al, and is always in the strong form, i.e. no weakening is indicated.
- b) No provision is made to indicate "hamzat al-wasl" or moon/sun letter distinction.
- c) The symbol \ after a letter indicates that that letter is elided in the writing system though it has a phonemic realisation. For instance, the alif in Al is elided when preceded by the preposition li. This is transliterated as "li A\l kitAbi".
- d) Ta' marbūṭa is represented by the letter (T).

App. 4.6 Resolving Problems of the Word

A detailed discussion on resolving the problem of the word in both English and Arabic is found in Appendix (1). This supplies the theorisation and pragmatic solutions for establishing the entity of word in the corpora.

It should be noted that Arabic words (according to our conception of the entity) that are orthographically amalgamated to another have been separated by a hyphen (with the exception of the connectives wa and fa, two connectives at the heart of this study; these are separated by a space). In the computerised scheme, such words are given full autonomy and are, therefore, separated by a space. For example:

Conventional scheme: bi-dirā<ay-hi

Computerised scheme: bi virAeay hi

Footnotes to Appendix 4

- (1) Wellisch (1978 p.13) defines a writing system as "a system of rules governing the recording of words and sentences of a language by means of conventional graphic signs". He adds (p.14) that the term is "entirely abstract and generic, and it is neutral as to function or typology". In addition, a writing system is always related to a particular language, such that a writing system of one language cannot be used for another. According to Wellisch, there may be similarities in the writing systems of two languages, foremost among them being the script used, but as systems "they are essentially independent of script, which is only one of their elements."
- (2) We have ignored in this brief outline those characters that have been added to the repertory of the Arabic script by non-Arabic languages that use it, such as Urdu, Persian, Kurdish, Sindhi, Swahili, Pushto.
- (3) This is an adaptation of Wellisch's (1978) classification of requirements.
- (4) Multiplicity of characters may be convenient in short transliterated texts, but may cause ambiguity when used for large chunks of text. For instance, Cantarino 1975 uses dh to represent ذ. But in his scheme both اذى [damage, harm] and اذهى [more cunning] have the same transliterated form "adhā".
- (5) In Oxford Computer Centre arrangements have already been made whereby automatic reproduction of a transliterated Arabic text in its original script is possible by using the Lasercomp.

Appendix (5A)

OCP Command File for Concordance Generation of English

```
*INPUT
TEXT COMMENTS "%%".
REFERENCES COCOA "(" TO ")".
SELECT EXCEPT BETWEEN "/" ".
*WORDS
ALPHABET "^ [ ] _ ' A=a B=b C=c D=d E=e F=f G=g
H=h I=i J=j K=k L=l M=m N=n O=o P=p Q=q R=r S=s T=t U=u V=v
W=w X=x Y=y Z=z & 0 1 2 3 4 5 6 7 8 9 $ # %".
DIACRITICS "/" "-".
PADDING "(" ) . : " + = !".
*ACTION
DO CONCORDANCE AND STATS.
REFERENCES X=3 WITH ".", L=5.
SORT KEYS BY START.

*FORMAT
PRINT USE "." AS "<", "," AS "|".
*GO
```

Appendix (5B)

OCP Command File for Concordance Generation of Arabic

```
*INPUT
TEXT COMMENTS "%%".
REFERENCES COCOA "(" TO ")".
SELECT EXCEPT BETWEEN "//".
*WORDS
ALPHABET "[ ] / < u a i # W N @ M Q=O=G=J ~ E Y
A=V=a`=ay`=>=u`=i`=An`=ayn`=aw` ' b t=T C j H
x d v r z s S c D p Z e g f q k l m n h w
U=uw]=u]=={Un]=I]=In] y I=iy]=i]=}=M] L 0 1 2 3 4
5 6 7 8 9 % | ``.
PADDING "\ : " ( ) ! +".
*ACTION
DO CONCORDANCE.
PICK HEADWORD "An" = "A[", "Un" = "UL[ U[",
"I" = "nI[ Ina ya", "lays" = "las[", "YilV" = "Yilay[",
"ealV" = "ealay[", "ladV" = "laday[", "kila" = "kilay[",
"kiltA" = "kiltay[", "EabU" = "EabA[ Eabi[",
"EaxU" = "EaxA[ EaxI[", "vU" = "vA[ vI[", "hum" = "him[",
"humA" = "himA[", "EuDpur#" = "EuDpurir[",
"EaHas#" = "EaHsas[", "Eacar#" = "Eacrar[",
"YiDpar#" = "YiDparar[", "EaHab#" = "EaHbab[",
"wad#" = "wadad[" OR SUFFIXES "u a i W N @ M [".
REFERENCES X=3 WITH ".", L=5.
SORT KEYS BY START.
MAXIMUM CONTEXT 50 LETTERS.
*FORMAT
PRINT EXCEPT "{ } [ >" AND USE "aw" AS "aw`",
"a" AS "a`", "ay" AS "ay`", "u" AS "u`", "i" AS "i`",
"u" AS "u]", "uw" AS "uw]", "i" AS "i]", "iy" AS "iy]",
"I" AS "I]", "ayn" AS "ayn`", "An" AS "An`", "Un" AS "Un]",
"In" AS "In]", "." AS "<", "," AS "|".
*GO
```


Appendix 6
Steps for Accessing OCP

App. 6.1 Preliminaries

The content of this Appendix is related to Chapter 5 (Vol. I) in the thesis. The aim is to outline the steps that have been followed in accessing OCP, particularly as far as processing Arabic is concerned. We shall discuss each main section of the relevant commands, outlining the options that have been selected and their intended effect.

App. 6.2 Input

Commands in this section define the format of the input text and what portion of it are to be selected for processing. We have used three types of commands for our major runs.

a. Comments: Words that are not to be included in the lists and concordances (such as vernacular expressions in the Arabic corpus) are treated as comments and included between "%%....%%".

b. References: According to the requirements of a COCOA format, references indicating the source of the text and its date of publication have been enclosed in double brackets "(". These are arranged in the input corpus as in the following example:

```
((X 219))  
((T Requiem for the Labour Party))  
((A Bernard Levin))  
((P The Times))  
((N 61559))  
((Y Wednesday))  
((D 15/6/1983))
```

(Code: X:serial number of text in the corpus; T:title; A:author; P:source newspaper; N:serial number of source issue as it appears on the paper; Y and D:day and date of publication).

References have to be declared, otherwise OCP will treat them as part of the text.

c. Select: All text subtitles have been retained. Since they are not to be included in the concordances, they have been enclosed in slashes "/" in the input text. This enables the command "Select except between //" to ignore them.

App. 6.3 Words

This section defines how the text is to be broken down into words. The following commands have been specified.

a. Alphabet: The following steps have been taken

- i. In English, the apostrophe is declared a letter and put at the start of the alphabet.
- ii. English upper and lower case characters are equated.
- iii. In Arabic, the original order of the alphabet has been preserved. However, the alphabet is preceded by the characters referring to the short vowels "u a i", the gemination character shadda "#" and "tanwīn" characters " W N @ M".
- iv. The long vowels "ū" (transliterated as U) and "ī" (transliterated as I) are placed after "w" and "y" respectively.
- v. The long vowel "ā", represented by "A", "V" and dagger 'alif are equated.⁽¹⁾
- vi. The different shapes of Arabic hamza are equated "Q = O = G = J". However the "madda", transliterated as "~", and hamza on alif, transliterated as "E", and hamza

under alif, transliterated as "Y", are given autonomous value in the alphabet.⁽²⁾

- vii. The ta' marbūṭa, transliterated as "T", is equated with "t": "t= T".
- viii. Numerals and the percentage sign are added to the alphabet of both English and Arabic. The pound and dollar signs are added to the English alphabet only; they are not used in the Arabic corpus.
- ix. To differentiate the decimal point from the full-stop, all decimal points in the input text have been replaced with the character "<", which is duly declared in the alphabet.
- x. While the general order of the English alphabet is fixed (see also Appendix 5), in Arabic it is set as follows:

"u a i W @ N M # Q=O=G=J E Y A=V=' b t=T C j H x
d v r z s S c D p Z e g f q k l m n h w=U y=I L
0 1 2 3 4 5 6 7 8 9 ".

xi. The silent "alif", transliterated as "L", is placed after the "ya'", just before the numerals.

b. Padding: The following are declared as padding characters
'\"()!+'.

c. Diacritics: The slash and hyphen "/-" are declared as diacritics.

For specifying other commands in the "Word" section, such as "Compress" and "Maximum word length", the default system is used.

It should be observed that the command and procedures discussed above are sufficient for picking and sorting the orthographic word. However, they are not compatible with the theorisation and procedures suggested in Chapter (5) and discussed in Appendix (1) for the definition and identification of word in Arabic. Some intervention is needed at this stage for adopting tagging techniques that may assist OCP in sorting key words according to the proposed specifications. The tags that will be used have to be declared in the alphabet. This point is treated in detail in Chapter (5).

App. 6.4 Action

This specifies the nature of the task that OCP performs, e.g. asking for word lists or concordances. We have limited the information related to the position of the key word in the text to the serial number of the text and the line number in the corpus. For concordances, the length of context that has been requested is 50 characters on each side of the key word. For other commands in this section, the default system is used. Other commands used for the Arabic corpus ("pick head word" and "pick suffixes") is discussed in Chapter 5.

App. 6.5 Format

This section specifies the format of the printed concordance. Except for the necessary change of "<", representing a decimal point, into ".", the default system is used. However, because of the intricate tagging procedures used for classifying words in

Arabic, more detailed specification of format is requested, whereby all tags (see Chapter 5) are to be ignored at the final printing of output.

Footnotes to Appendix 6

- (1) It is not necessary to equate the dagger 'alif with other forms of the same letter. This is because its use is highly restricted and cannot cause confusion in sorting.
- (2) This seemingly arbitrary procedure is justified on the grounds that a) these forms of the hamza represent three different phonemic combinations, which explains why they are given different graphemic shapes in the Arabic writing systems; b) unlike the other graphemic shapes of the hamza, the madda, hamza on 'alif and under 'alif may occur as word initials, and, therefore, placing them in a particular order is necessary.

Page removed for copyright restrictions.

Appendix 7C

A Sample Page with Text Deleted
and Connectives Retained

_____ also _____
_____, but _____
_____.

_____,
_____ provided _____
_____ for
_____, however,
_____.

_____ when _____
_____ for _____

_____.
_____.

In the first instance, _____
and on that occasion _____
_____.

_____.
_____.
_____ if _____
_____ and _____
_____.

_____ further _____
_____ where _____
_____, or _____
_____. On the other hand _____
_____, as _____
_____.
_____.