Computer-Aided

Categorisation and Quantification of Connectives

in English and Arabic

(Based on Newspaper Text Corpora)

Volume (2)

Adnan Jabbar Radhi Al-Jubouri

Doctor of Philosophy

The University of Aston in Birmingham

July 1987

PREFACE TO VOLUME (2)

This volume comprises Chapters 6, 7, and 8, which proffer a detailed analysis of the textual (both qualitative and quantitative) properties of connectives observed in the English and Arabic corpora. More specifically, the three chapters achieve a two-fold task: a) description of the textual behaviour of connectives in each corpus, with particular emphasis on their their cohesive function and organisational role; b) identification of the quantitative patterns of the textual operationality of connectives in the text.

The plan of the Volume is as follows. Chapter 6 offers a detailed account of the functioning of connectives, with exemplification drawn from both corpora. Chapter 7 achieves two major tasks. First, it introduces a statistical profile of each text corpus, discussing the frequency distribution, patterns of repetitiveness, and growth of types and their relation to the size of text. This is intended to serve as a background for the second, more fundamental task: statement of a general calculus of observations on the properties of connectives. Chapter 8 is concerned with outlining a quantitative profile of each of the various functional categories identified in the corpora.

Interlingual contrastive statements have been kept to a minimum. The main aim at this stage is to acquire an understanding of connectivity before any contrastive effort is initiated. This is in conformity of the investigatory plan suggested and elaborated in Chapter 2 of Volume (1). A detailed differential account is provided in the next volume.

CONTENTS

PAGE

PAGE

PAGE

LIST OF TABLES

LIST OF FIGURES

PAGE

CHAPTER SIX

Categorisation and Description of Connectives

6.0 Perspective

One of the central tasks of this study is the provision of a textual description of connectives in English and Arabic that stresses their cohesive nature and patterns of functioning, and that may help to outline interlingual textual variations. This description will be based on evidence from the two bodies of natural texts that have been automated and analysed. Any possibilities of textual functioning that are not substantiated by reference to the corpora but may be obtainable on the grounds of intuition will be disregarded. The aim, as has been argued earlier, is to derive our textual account of connectives from empirical observation of samples of language in use. This means that all exemplificatory extracts are taken from the corpora.

The general aim of this chapter is to provide this description. It intends to achieve this major task by dividing it into phases:

1. Specification of the structural patterns of connectives.
2. Functional categorisation of connectives: description of scheme and categories.
3. Description of the cohesive functionality and textual patterning of connectives.

To achieve the first phase, we start with an introduction that reviews various terms used in linguistic studies for labelling connectives, pointing out the linguistic dimension of each label and

justifying our own terminology. We then describe the structural patterns of the items that we are labelling connectives in this study.

The second phase is approached by describing the scheme of categorisation: its purpose, criteria and operationality. This is followed by a specification of the functional categories as located in the corpora.

The third phase is the main one in this chapter and by far the most comprehensive. Each main category of connectives is first examined from a broad perspective and its main characterising features of functionality are discussed. A subcategorisation is then made into types that exhibit sufficiently distinctive functioning in the text. In each subcategory the textual repertoire of connectives is mentioned, and later its specific cohesive functions and textual patterns are discussed with ample exemplification. Where textual patterns are not clearly distinctive, emphasis is placed on functionality, and discussion of textual patterning is dropped.

Each illustrative excerpt used in the discussion of functionality is followed by its reference: its source, date of publication, number of text in the order of texts within the corpora, line numbers within the corpora. For example, a reference such as "G, 11/4/83, X24, 3576-86" means the excerpt is taken from the Guardian of 11 April 1983; text number in the corpus is (24) and the line numbers that represent the place of the excerpt in the corpus are 3576-3586. Code letters for the source of the excerpts (names of the newspapers) are self-explanatory, but will be mentioned here for

clarity.

G	The Guardian
O	The Observer
DTel	The Daily Telegraph
STel	The Sunday Telegraph
Ti	The Times
STi	The Sunday Times

Code for Arabic text sources:

Nb	Al-Anba'
Hr	Al-Ahram
Th	Al-Thawra
J	Al-Jumhuriyya
Sh	Al-Sharq Al-Awsat
Ar	Al-Arab

In Arabic each excerpt is produced in a transliterated form using the conventional scheme discussed in Appendix (4). This is followed by a translation of the excerpt into English. The translation is intended as an approximation to the original Arabic text since the aim is to assist the reader to gain a better understanding of the examples. We admit that more experienced translators may produce a slightly different rendering of the same excerpts.

A note of warning is in order. For pragmatic reasons the description of English connectives and that of Arabic connectives are merged into one that exploits general functional features. This should by no means imply that English and Arabic connectives have identical functional patterning. At this stage we are not interested in pointing out variations, and any contrastive effort is incidental. As we have agreed in earlier chapters (particularly Ch.

2) the investigatory apparatus of the textual contrastive analysis requires, for efficient operationality, a comprehensive description of the phenomenon we set out to study. An elaborate contrastive account, where variations both qualitative and quantitative are discussed, will be postponed till later (see Ch. 9).

Finally, the description in this chapter is sufficiently detailed to suit the purposes of this study. A more comprehensive analysis is feasible but will then burden the thesis beyond reasonable limits. Therefore, although the description is still open for more contribution, the present one is, within the restriction of this project, felt to be satisfactory.

6.1 Problem of Nomenclature

The disagreement in the conception of a class called connectives and the conflict in views regarding their syntactic and semantic function has led to the adoption of a variety of terms to label the same class. This problem was pointed out as early as 1952 when Fries explained that one difficulty with the description of conjunction is defining what a conjunction is and which words are conjunctions. He observes that even in the "American College Dictionary", yet, defined as "nevertheless", is labelled as a conjunction, while nevertheless, defined as "however", is labelled as an adverb, and however, defined as "nevertheless" or "yet", is labelled as a conjunction (Fries 1952, p.250).

Confusion can be traced to earlier grammarians. We have already noted how Sweet (1892) classifies the items that express clause connection. A similar classification is made by other later linguists. For instance, Whitehall (1951) recognises two types of

conjunction: one covers coordinators and subordinators while the other covers "conjunctive adverbs" which in his view operate like other conjunctions but differ in that they (a) carry stress in speech, (b) can introduce paragraphs and sentences which are not in association, (c) can appear anywhere in a sentence and (d) are preceded by a semi-colon in initial sentence position (ibid p.70). This label, i.e. "conjunctive adverbs", is rejected by Hill (1958) who, due to the emphasis he places on structure (compared to function), prefers to call them adverbs. In his structuralist arguments, Hill narrows the class of conjunctions to and, but, or, as, if, because, till, until, although and unless (Ibid p.402). A position that combines both Whitehall's and Hill's views is taken by Chatman (1964) who maintains Whitehall's distinction but Hill's stance that conjunctive adverbs are more "adverbial" than "conjunctive".

Francis (1958 p.415) talks about "sentence modifiers" or "sequence signals" when he refers to both conjunctive adverbs and such prepositional phrases as "on the other hand" and "in contrast". Bolinger (1965 p.289) prefers to call them "conjunctive adverbs" while Greenbaum (1969), in his detailed study of adverbials in English, calls them "conjuncts". Greenbaum's classification is echoed in Quirk et al. (1972) and (1985) (see 4.4.2 above).

Conjunctive adverbials are often called "sentence connectors" (for instance, Quirk et al. 1972, 1985, Wijasuriya 1971, Frankel 1977, Chatman 1964). This is a restricted term and covers only one group of connectives. Some stylists use the label "transitions" (for instance, Winterowd 1975), a reminiscence of the use of the

same term in phonology. In cohesion models, Beaugrande and Dressler (1981) call them "junctives" while Halliday and Hasan (1976) label them "conjunctive devices" or "conjunctions". Some psycholinguistic studies prefer to use the term "logical connectives" (for instance, Gardner 1977) following the treatment of connectives in some semantic studies, though most psycholinguistic work prefers to use "connectives" (for instance, Robertson 1968, Paris 1973, 1975, Beilin and Lust 1975, Bloom et al. 1980).

In this study, we shall use the term "connectives". The choice of this term is justified on the grounds that terms such as "conjunctive adverbs" or "conjunction" are limited in scope, restricting the study to a small number of items. Both, in addition, give a syntactic slant unnecessary in this study; our main interest lies in the cohesive role of the connectives and therefore their syntactic properties will not be discussed except where such a discussion is deemed desirable. In Arabic, the use of these two terms is even more restrictive than in English: first Arabic has a limited category of adverbs and secondly conjunction (as we have seen in 4.1.2 above) traditionally comprises a small number of coordinators; it does not even include subordinators. Moreover, the use of the term "conjunction", whether in English or Arabic, is misleading. On the one hand, it is confused with a particular type of connectives in logic, denoted by the operator [•]. On the other, it will suggest inclusion of "phrasal connectives", i.e. conjunctions that join two terms in complex terms of one grammatical category (e.g. two nouns, two adjectives, two adverbs etc.). In this study such function of the connectives is excluded since, as has been discussed in 4.4.1.2, the connected units are

"propositions" referring to the content of clauses and clause-complexes. The use of the term "conjunction" will also allow two types of subordinators which will also be excluded in this study: (a) the complementiser "that", and also those interrogative particles that are used in reported questions; (b) relative pronouns and adverbs. The main reason for not including these two types is that their function is, in our view, fundamentally syntactic: type (a) introduce the subject, object or complement and type (b) provides an expansion for the noun phrase (except in the case where "which" is used in a non-restrictive way referring to a whole proposition rather than a single noun).

Other terms used in the literature are also not convenient for our purposes. The term "conjunctive device" is awkward and cumbersome; the term "transition" is vague and unfamiliar; and the qualification in "logical connectives" gives the impression that the study is concerned with connectives in a logical or formal logico-semantic framework. Additionally, the structural types of the connectives as we envisage them in this study are broader than that subsumed by any of the labels mentioned above. This point will be made clearer in the next section.

6.2 Structural Types of Connective Expressions

Before we embark on a categorisation of the textual relations signalled by connectives (the main task of this Chapter), we would like to identify the types of items that are termed "connectives" in this project. The arguments for using the term "connectives" (6.1.1 above) and the discussion of the textual characteristics of the connective in 4.5 have made it clear that textual relations are not

uniquely signalled by conjunctions, nor by the items that are frequently classified as conjuncts. These relations may, in fact, be explicitly signalled by a range of structures. In the corpora we have identified the following:

1. Conjunctions: The conventional subclassification of this group is coordinators and subordinators. In English the distinction is best discussed in Quirk et al. (1972, 1985). Coordinators include and, but and or. Subordinators include a variety of expressions: when, while, because, if, so that, etc. In Arabic, coordination and subordination is comprehensively discussed in Cantarino (1975 Vol.III). Arabic coordinators include "wa", "fa", "'aw", etc.; subordinators include such items as "'iḏā", "<indamā", "'in", etc.

2. Conjuncts: This refers to a range of adverbials that have the function of conjoining independent units, here clauses or sentences. Such adverbials include however, in addition, accordingly, thus, in other words, for example, etc. English conjuncts have been described syntactically under adverbs or conjunctive adverbs or adverbials (see references in the last section), and their semantics have also been outlined (cf. Greenbaum 1969, Quirk et al. 1972, 1985, Quirk and Greenbaum 1973). But in Arabic a class of conjuncts have rarely been discussed although it does exist. In traditional grammar, in particular, one-word conjuncts have mainly been covered by the term "adverbs" [ẓurūf] or "particles" [ḥurūf], interest being fundamentally on their impact on the morphological status of neighbouring constituents of the sentence in which they occur. For the purposes of the analysis,

Arabic conjuncts have been identified, but included within the class of connectives. No attempt has been made to distinguish them from other connectives on the basis of their syntactic function, as, in our view, such an attempt is worth a separate study.

3. Conjunctive phrases with an anaphoric reference: This refers to a class of multi-word conjuncts that have cohesive power because of the occurrence of an anaphoric reference. Whether such instances should be considered as conjunctive or as reference is speculative and disputable. Strictly speaking, according to Halliday and Hasan (1976 p.230),

> "they belong with reference, because they depend on the presence of a reference item... But since they involve relations which also function cohesively when expressed WITHOUT the accompaniment of reference items, it is simpler to include them within the general heading of conjunction" (their emphasis).

Examples for such connectives: because of that, instead of this, since then, etc. Examples from Arabic: "li-hāḏā al-sababi" [for this reason], "badalan min ḏālika" [instead of that], "bi-al-'iḍāfati 'ilā hāḏā" [in addition to this]. The demonstrative element in the phrases (i.e. the reference item) usually refers back to all the previous proposition or to a sequence of prepositions, exhibiting an enormous cohesive power. However, it may occasionally refer back to a single item, usually a point of focus in the previous proposition. (Exemplification will be offered later when such connectives are examined for their cohesive force.)

4. Subjuncts: Subjuncts are a class of adverbials that differ from adjuncts in respect to a number of syntactic and semantic features (for a discussion of these features see Quirk et al. 1985).

For instance, adverbs such as economically may either operate in the semantic role of "process", where it is treated as an adjunct, or in the semantic role of "respect" or "orientation" where it is treated as a subjunct. Occasionally, subjuncts may have a role that, while it applies to the whole clause in which it occurs, creates specific modification or establishes a different angle of orientation for the subsequent stretch of text in comparison to the current one, thus achieving a cohesive effect. Subjuncts that have this role are here considered connectives, e.g. economically speaking, politically, certainly, etc. In Arabic subjuncts comprise a very limited group of expressions, e.g. "ḥaqqan", "fi<lan", "xāṣṣatan", etc.

5. Disjuncts: Disjuncts share with conjuncts similar syntactic properties. Compared to the other elements of the clause, both are more "syntactically detached" and in some respects "superordinate" in that they "seem to have a scope that extends over the sentence as a whole" (Quirk et al. 1985 p.613). However, they exhibit differences in their semantic role. While conjuncts have the function of conjoining independent units, disjuncts seem to contribute an additional facet of information to a single integrated unit: relating it to the text producer's "authority" or "stance".

It should be noted that we have exercised considerable caution in including disjuncts as connectives in this study. We could have left the door ajar to permit all disjuncts; but then certain disjunctive elements would have been hard to justify as connectives in the way we have characterised them (as in 4.5 above) (cf., though, Knowles 1984, where a variety of disjuncts in Polish are classified as a category of Particles). Accordingly, only

disjunctive expressions that textually relate a proposition, in a certain way, to a previous one or more have been allowed in the analysis. This task is not easy to perform automatically (unless considerable effort is expended on establishing semantic networks that require sophisticated programming, a task that is beyond the scope of this project) and therefore has in this case to be achieved manually.

Disjuncts are divided, in Quirk et al.'s 1985 terminology, into two subclasses: style and content disjuncts. The former (i.e. style disjuncts) conveys the text producer's comment as to a) modality and manner (e.g. honestly, truly, seriously); and b) respect (e.g. in broad terms, figuratively, literally) (ibid pp.615 ff). Content disjuncts, also known as "attitudinal disjuncts", particularly in Greenbaum (1969), Quirk et al. (1972), Quirk and Greenbaum (1973), make an observation concerning a) degree of conditions for truth of content (e.g. admittedly, certainly, allegedly), b) value judgement of content (e.g. unwisely, rightly, obviously, certainly) (see Quirk et al. 1985 pp.620 ff).

6. Adverbial phrases with an embedded relative clause: This is a very restricted class, and includes mainly those adverbial phrases that are basically time or space relators, but which contain a relative clause introduced by the relative adverbs when or where (e.g. at a time when, at a point where) or, occasionally, by relative pronouns (that/which in English, "alladī/allatī" in Arabic). The semantics of such constructions are similar in some respect to time or place subordinate clauses and it is felt that they should similarly be included in the analysis of connectives.

7. Non-finite verbal expressions: This refers to the use, in

English, of non-finite verbal phrases, particularly infinitive, to relate a clause/clause-complex to a previous text component, e.g. to conclude, to put it differently, etc. Arabic is marked by absence of similar connectives.

8. Clausal expressions: This is a limited set of connectives, each is composed of a clause with a finite verb form. When used on their own, these clauses barely have a conjunctive value. But the way they are structured and the manner in which they are strung with the text sequence make them function as textual connectives. As with disjuncts, we have been careful in the identification of this class of connectives. Only those that have clear connective function have been labelled and included, e.g. in English: Add to that; in Arabic "'aḍif 'ilā ḏālika". Ambiguous cases have been left out. For instance, "performatives" (Austin 1962), such as I think, I assume, I realise and I doubt, though accepted as connectives by Werth (1984), are not included here.

6.3 Categorisation of Connectives

6.3.1 Features

As mentioned earlier, there have been a number of schemes for categorising the semantics of connectives. We would like now to show how the categorisation in this project differs from those conceived. We shall restrict our discussion to the features of the "process" of categorisation.

1. The categorisation is based on an examination of the role of the connectives in signalling textual relations that bind one proposition or a sequence of propositions with another. More specifically, it is an examination of the devices used for

connecting the next block, or blocks, of conceptual text-world knowledge with the current one(s), in such a way that the various relations among knowledge configurations are made explicit. Hence, certain categories, totally ignored in some categorisation schemes, are here identified and their textual nature discussed.

2. The categorisation is applicable to connectives in both English and Arabic and is intended to establish the extent of the variation in the way text in each language is organised. This task dictates that the process of classifying a particular relation is to be flexible enough as to be able to create a spectrum that encompasses minute semantic variations. Hence categorisation is not treated as a strictly discrete compartmentalisation of relations. This point will be elucidated further as the categories are described.

3. The categorisation is based on an examination of the behaviour of the connectives in two corpora of natural texts. No exemplification has been contrived in the description of the categories, unless, of course, special reference is made to the work, including examples, of other investigators and scholars. We hold the view that such a categorisation offers a better approximation to reality and provides a stronger empirical evidence for the typology of textual relations signalled by connectives (see discussion in Ch. 5).

4. Certain stages in the categorisation are computer-aided. More specifically, computer techniques have been used to identify connectives in the corpora (see Ch. 5); this has insured that candidate expressions have been isolated and some broadly

categorised. However, in detailed categorisation of textual relations, manual intervention is inevitable.

5. Categorisation in this project follows a bottom-up approach. That is, detailed subcategories are first identified in the corpora and the connectives are duly tagged. Next, these subcategories are grouped according to common textual properties into larger categories, which has finally given us the categorisation scheme. Unlike some other categorisations, we have not designed a scheme with superordinate categories, classified into subcategories and then imposed on the data. This, if it had been followed, would have gone contrary to the aims and methodology of the project, and would have stripped it of its empirical value.

6. The schemes that have been conceived for categorising connectives fall into two main groups according to complexity. The first comprises a small number of categories with a large number of subcategorial distinctions. To this group belongs the scheme set-up within the functional systemic model of cohesion (Halliday and Hasan 1976, and their followers, for instance Martin 1977, 1983): five major categories of conjunctive cohesion, but with detailed subcategorisation. The second main group of schemes offers a large number of main categories without, or with limited, subcategorisation. To this group belongs the schemes adopted by Greenbaum (1969), Milic (1969) and Quirk et al. (1972, 1985). Each group has its theoretical as well as practical advantages and disadvantages. The question is a trade-off between the clarity of the logico-semantic description and the practicalities of the analysis. We would like to adopt a compromise in the hope of achieving a clearer understanding of connectives. We have,

therefore, grouped the subcategories into a large list of nine categorial classes. Further grouping is also attempted, but on different principles and will therefore be discussed in a later chapter.

6.3.2 Composition of the scheme

Textual relations signalled by connectives are first grouped into 28 categories. Next, they are then grouped into nine main categories. The composition is outlined in Table (6.1):

	Category	Tag used	Main Category
1	Appending	XAa	
2	Enumeration	XAn	Additive
3	Amplification	XAm	
4	Comment	XAo	
5	Continuity	XAc	
6	Coupling	XAt	
7	Similarity	XCs	
8	Degree	XCd	Comparative
9	Alternation	VRl	Alternative
10	Restatement	LFr	Reformulatory
11	Exemplification	LFx	
12	Summary	LFm	
13	Adjustment	LMv	
14	Confirmation	LMf	Orientative
15	Sequence	NTq	Temporal
16	Simultaneity	NTm	
17	Span	NTs	
18	Temporal Point	NTp	
19	Temporal Circumstance	NTc	
20	Time Frequency	NTf	
21	Space	NSr	Spatial
22	Cause/Reason	NCs	
23	Result/Inference	NCn	
24	Magnitude/Degree	NCd	Causal
25	Purpose	NCp	
26	Condition	NCc	

27	Antithetic	VDc	Adversative
28	Contrast	VDd	

Table 6.1 Categorisation Scheme

In discussing these categories, we shall pick each main category and discuss the subcategories associated with it. This is a convenient way to control the whole account. A description of each subcategory will be given for both English and Arabic. The contrastive statement is delayed until a full descriptive profile is provided in both languages (see Ch.2 for a full analysis of the method).

6.4 Additive

6.4.1 General Comments of Textual Role

This is a basic type of connection. The usual relation is that of listing one block of text-world knowledge after another. That is, the next proposition is seen as an addition to the current text sequence.

Connection through additivity requires, among other things, that the knowledge configurations that are linked are similar (for instance, two, or more, events, situations, states, etc.), or from the same semantic domain. This is the customary textual pattern and can be exemplified as follows:

[6.1] His wife is an invalid and his present salary is about one-twentieth of what he could expect to earn.
(DTel, 28/3/83,X100,16885-7)

The relation expressed by and is that of adding or, rather, listing another "worry" or unfavourable state.

However, the exigencies of text structure may bring together unusual items in ad hoc domains relative to the text-world

knowledge. The pattern in this case is "novel" (cf. Longacre 1983). For instance, in this example:

[6.2] Carrots were off the menu and the need to get rid of surface labour was too great for much finesse.
(DTE1, 1/6/83, X114, 19066-8)

the second proposition is from a semantic domain that is different from that of the first. The first proposition is a continuation of an image set in the previous text sequence (The manager ruled with the stick and the carrot.) Therefore, despite the novel textual pattern, the second proposition is still an addition to the first one; both describe the resultant state of a particular situation.

Although "pure" addition is non-temporal and non-causal, it is often difficult to strip the additive relation from temporality and/or causality. However, we have been careful in distinguishing degrees of additivity, temporality and causality. Where temporality or causality is a stronger element in the relation, we have considered the connective temporal or causal respectively. This explains the identification and tagging of a temporal and causal "and" (in the English corpus) and "wa" (in the Arabic corpus).

6.4.2 Additive Categories

In the two corpora, additive relations that are signalled by connectives can be categorised into six subcategories: 1) appending 2) enumeration, 3) amplification, 4) comment, 5) continuity, 6) coupling. All types denote relations among propositions, the fifth one, "continuity", however, includes, a function of the connective relevant to its environment. Each of these categories expresses a different aspect of additivity. That is to say, although all relations create text via a process of addition, each directs the

process in a different path which, eventually, arrives at a different fitting for the next block of knowledge. Hence each type of relation organises text differently.

It is worth mentioning that these categories can be further subcategorised into various more subtle relations, and connectives can be examined at a further level of delicacy. However, for the purposes of the analysis in this project, and in order to keep the description within manageable limits, we shall restrict our discussion to this level. These categories will be examined in detail next.

6.4.3 Appending

6.4.3.1 Repertoire

Connectives identified as signalling the additive relation of appending in the corpora comprise the following two sets:

A. English

Simple (one-word): additionally, again, also, and, besides, either, further, furthermore, moreover, neither, nor, then, too.

Compound (multi-word): add to that, added to that, as well, as well as, in addition, more than that, what is more.

Correlate: neither ... nor

B. Arabic

Simple: wa, kamā, 'ayḍan, kaḏālika, ṯumma , fa, 'aw, mujaddadan

Compound: min jadīdin, marratan 'uxrā, bi-al-'iḍāfati, yuḍāfu 'ilā, faḍlan <an, 'ilā jānibi,

'iḍāfatan 'ilā, fawqa (ḏālika, hāḏā), yuḍāfu li, al-'akṯaru min hāḏā, ṯāniyatan, 'aḍif 'ilā, 'akṯaru min ḏālika, marratan ṯāniyatan, bi-jānibi 'anna, min ṯamma, min jihatin 'uxrā, wa 'akṯaru min .

6.4.3.2 Textual Functioning

This relation is additive proper. In both languages, appending connectives have two roles achieved simultaneously:

a) They convey an incremental effect where an event or a state (the subsequent or next connect) is annexed to the current one (the antecedent). Since the two events or states are usually related (e.g. from a similar conceptual domain), it follows that their combination provides a reinforcement that functions in two directions: from antecedent to subsequent, though mainly from subsequent to antecedent. For example:

[6.3] Every penny paid out this year must be paid in this year and it is paid in by the working population.
(Ti, 22/11/82,X167, 27327-29)

[6.4] ... Al-niqasu hawla-ha lam yatawaqqaf wa la 'azunnu-hu sawfa yasilu 'ila nihayatin fi al-mustaqbali al-qaribi, 'aw hatta al-ba<idi.
[... arguments about it (Camp David treaty) have not ceased and I do not think they will come to an end in the near future.]
(Hr, 30/3/83, X41, 7525-7)

b) They convey a sense of ordering of propositions that denote related blocks of knowledge. Often the ordering follows the normal ordering of the facts themselves across temporal, spatial or causal axes, although there can often be no "internal" (i.e. essential) relation of temporality or cause and effect. Examine these examples

in English and Arabic:

[6.5] What we did to the poor of the earth has come home. And it has brought to us the same disease that was inflicted on the Third World.
(G, 17/1/83, X13, 1989-90)

The connective "and" appends the two propositions and simultaneously signals an ordering (both temporal and causal) of the two sets of facts. Similarly in Arabic:

[6.6] ...Pakistan maṯalan allatī kānat tastawridu 3 malāyīni ṭannin min al-qamḥi fī sanati 1978 ḥaqqaqat hāḏā al-<āma al-'iktifa'a al-ḏātiyya fī maḥṣūli al-qamḥi, wa tuwaffiru fā'iḍan li-al-taṣdīri...
[Pakistan has achieved this year self-sufficiency in grain and has got a surplus for export ..]
(Hr, 15/4/83, X48, 8946-9)

the connective "wa" appends the two propositions additively but marks a clear ordering.

In addition to this normal ordering, there may be another ordering of propositions, determined by the requirement of relative interpretation. For instance, a particular way of appending is used in order to adjust the perspective through which a set of conjoined blocks of knowledge are to be viewed. This is exemplified in the following two excerpts.

[6.7] Labour is hooked on a political ideology that the national majority refuses to accept. Moreover, the non-Labour votes are very much more positively anti-Labour than the non-Conservative total of (Labour and Alliance) votes is anti-Conservatives.
(STi, 12/683, X250, 39259-63)

[6.8] alladī yajrī fī būlandā al'āna šay'un garībun wa <ajībun wa muhayyirun li-al-'aḏhāni. wa yabdū l-ī 'aḥyānan 'anna al-suluṭāti al-būlandiyyata la-hā yaḍun fī ṯawrati al-jamāhīri; faqaṭ lā tastaṭī<u 'an tujāhira ḥattā lā taktasiḥa-hā al-dabbābātu al-sūfītiyyatu.
[What is happening in Poland now is strange and puzzling. And it appears to me sometimes that the Polish authorities may be giving a helping hand to the people's revolution (solidarity); but they

cannot announce this publicly in case they be invaded by Russian tanks.]
(Ar, 24/6/83, X216, 36923-8)

In each excerpt, the antecedent establishes a topic or point of departure. The perspective is expanded through the use of the connective moreover in [6.7] and "wa" in [6.8].

A related textual function which is pertinent to additivity is the way it creates and sustains parallelism. Here form and meaning collaborate to produce a rhetorical effect, an aspect that is of particular textual importance in Arabic argumentative discourse (see Chapters 9 and 10 for more details, see also Al-Jubouri 1984).

[6.9] kamā yusajjilu al-mahāwiya wa al-xuṭūba li-yanfuṭa fī al-'ummati rūḥa al-yaqẓati, wa yastanhiḍa min-hā xāmida al-himami, wa yuḥyiya fī-hā mayyita al-<azā'imi.
[In addition, [poetry] records crises and catastrophes so as to blow in the nation the spirit of awareness, and awaken sleeping capabilities, and enliven dead wills.]
(Hr, 4/1/83, X119, 4932-5)

[6.10] ...lam tajma<-ā šamlan, wa lam tuḥaqqiq-ā ḥallan, wa lam tubaddil-ā bi-salāmin ḥarban, wa lā bi-'ittifāqin xilāfan xilāla <umri-himā al-ṭawīli...
[... they [the Arab League and the Islamic Conference Organisations] have never united a nation, nor have they achieved any solution, nor replaced peace for war or concord for discord during their long period of existence ..]
(Sh, 18/3/83, X119, 23022-4)

Note that in both examples the parallelistic forms are all connected via "wa". They are well-spaced and sustain a particular rhythm (and in [6.10] intensify this rhythmic repetition by having rhymed endings).

6.4.3.3 Some Textual Patterns

In connecting blocks of knowledge, additive connectives of appending displays a number of patterns. The main ones are:

a) Repetition: Some connectives express a repetition of an earlier block of knowledge. The repeated block is configured in such a way that it yields a further addition, having its own control centre. In information processing, the existence of the connective helps verify and support the repetitive nature of the subsequent block of knowledge. To this group belong the connectives: English: again, also, too, again, Arabic: "'ayḍan", "wa".

[6.11] If the call came again, though, one suspects that Mr. Steel could whip them in.
(DTel, 1/6/83, X113, 18952-4)

[6.12] Glasgow, too, ... is nowadays practically denuded of ships.
(Ti, 21/1/83, X173, 28121-3)

[6.13] wa qad <amma al-sukūtu fī al-'āwinati al-'axīrati kaffata al-'awsāṭi. fa 'isrā'īlu sākitatun..
wa al-'amīrkiyyūna sākitūna ..
wa al-'ūrubbiyyūna sākitūna ..
wa al-rūsu 'ayḍan sākitūna ..
[Silence has lately dominated all parties concerned. Israel is silent ... And the Americans are silent ... And the Europeans are silent ... And the Russians are also silent ...]
(Ar. 30/6/83, X219, 37094-117)

Each makes a repetition of a statement made earlier in the text. Notice that in Arabic the connective "'ayḍan" [too, also] is often supported by "wa" or "kamā" [besides, in addition]. These two usually introduce the statement, i.e. occur in initial position, while "'ayḍan" normally occur in medial, though occasionally final, position to reinforce the additive relation internally (from within the statement itself).

b) Addition of a related block of knowledge. This can take one of two possibilities.

i) The connective may additively connect one concept in a

subsequent with the whole knowledge configuration expressed in the antecedent. For example:

[6.14] It is not merely that Francois Mitterrand is more resolute over strengthening his national nuclear deterrant than any French leader since de Gaulle ... Mitterrand is also an idealist ...
(STel, 29/5/83, X136, 22788-802)

[6.15] 'inna qaḍiyyata Lubnāna tamassu bi-šaklin mubāširin qaḍiyyata filasṭīna, wa 'ayḍan qaḍiyyata Sūriyya, wa kaḏālika qaḍiyyata Al-'urduni wa qaḍiyyata Miṣra, kamā tamassu 'ayḍan duwala gayri al-muwājahati ḥasaba al-taqsīmi al-jugrāfiyyi..
[The Lebanese crisis concerns [is directly related to] the Palestinian crisis, as well as the Syrian, Jordanian and Egyptian crises; in addition it also concerns non-neighbouring Arab countries according to their geographical position on the map]
(Ar, 30/6/83, X220, 37345-50)

In [6.13], a quality expressed in the consequent statement is added via the use of "also" to the content of the antecedent. Similarly in [6.14], another group of Arab countries, i.e. "non-neighbouring", is added to the groups mentioned in the current statement.

ii) The connective can additively annex two contextually related propositions, each with its own knowledge configuration. For instance, the two propositions in each of these two excerpts, though different in content, are contextually related, and textually connected by "and" and "wa" respectively.

[6.16] This is the great Tory hope for the economy, and the omens are good.
(STel, 12/6/83, X145, 24335-6)

[6.17] 'ixtalaf-ū; wa kāna xilāfu-hum bi-al-ḥujjati wa laysa bi-'ilqā'i al-tuhami.
[They disagreed; and their disagreement was based on proof (and evidence), not on (groundless) accusations].
(Sh, 15/2/83, X109, 21278-9)

c) The connective may do function b.ii but at the same time signals a shift in the participants, event or conceptual domain,

from one sentence to the other, and yet the two sentences are part of the text. This is illustrated in these two excerpts:

[6.18] [Cloning] might lead to an accumulation of recessive genes and an imbalance in the sexes. And who would decide who would be cloned?
(G, 2/12/82, X6, 836-9)

[6.19] wa la yumkinu 'an tazdahira al-majallātu al-ḥukūmiyyatu 'illā 'iḏā kāna 'iṣdāru al-majallāti ḥaqqan mutāḥan li-al-jami<i. wa naḥnu narjū li-hāḏihi al-majallāti kulli-hā 'an tanjaḥa wa tazdahira.
[State-published journals will not develop unless publication of journals becomes a common right for the public. And we wish all these journals success and expansion.]
(Sh, 22/2/83, X111, 21579-83)

In such instances the shift is soft or smooth and the relation is still clearly additive. In cases where the shift is sudden or prominent, the function of "and", though still basically additive, is predominantly adversative. This will be discussed later.

6.4.4 Enumeration

6.4.4.1 Repertoire

The following two sets comprise the additive connectives of enumeration located in the two corpora.

A. English

Simple: first, firstly, initially, and, second, secondly, third, thirdly, then, ultimately, finally, last, lastly.

Compound: first of all, for a start, to begin, to start.

B. Arabic

Simple: wa, 'awwalan, ṯāniyan, ṯāliṯan, 'axīran, rābi<an, xāmisan, sādisan, bidāyatan, sābi<an, ṯāminan,

tāsi<an, <āširan.

Compound: fī al-nihāyati, bādi'a ḍī bad'in, fī al-xitāmi, ḥādi <ašara, 'awwala mā bada'a, fī al-bidāyati, fī al-marrati al-ṯāniyati, fī al-marrati al-'ūla.

6.4.4.2 Textual Functioning

Additive relations, we have mentioned, essentially involve a listing of related propositions. Hence, all additive connectives have an enumerative function by default. However, enumerative connectives play a specific role in organising text by denoting an explicit cataloguing of an inventory and marking the items in a particular sequence.

The listed propositions can follow one another so that enumeration is intense, or they can be spaced throughout the text, each item filling a sizeable sequence of text, such as a paragraph or even spreading in a series of paragraphs. On a conceptual level, the listed propositions should have clear content relations so that each proposition (or set of propositions) represents an additional aspect, phase or degree. The following excerpts exemplify the function of enumeration in English and Arabic.

[6.20] The party [in Poland] has a number of problems. First it is the unchangeable party of government and must therefore be held responsible for everything that makes life miserable for young people - the 15-year wait for a flat for young married couples, the limited travel possibilities, the expense of food, the scarcity of cars, the poor quality of baby food.

Secondly, solidarity clearly identified party corruption as its main target. The villas and perks of the party leadership were contrasted with the lifestyle of the ordinary 25-year-old.

Finally, the church under martial law has continually laid claim to being the true representative of the nation, its voice at a time when no other form of political opposition is

practical. (Ti, 6/4/83, X188, 30340-56)

[6.21] wa la<all-ī 'antahizu furṣata 'iqtirābi safari al-ra'īsi Ḥusnī Mubārak 'ilā al-wilāyāti al-muttaḥidati fa 'u<ālija mā sabaqa 'an ta<arraḍ-tu la-hu fī hāḏā al-maqāli ṭāriḥan taṣawwur-ī li-mā 'antaẓiru-hu min xuṭuwāti al-siyāsati al-miṣriyyati al-muqbilati bi-ša'ni qaḍiyyati-nā al-filasṭīniyyati.

'awwalan 'ijhāḍu al-di<āyati al-mugriḍati fī al-mujtama<i al-'amīrkiyyi allatī tuḥāwilu tašwīha jiddiyyati mawqifi miṣra min al-salāmi allaḏī 'aṣbaḥa manhajan <arabiyyan ba<da qimmati fās.

ṯāniyan: 'istiṯmāru al-'āṯāri al-'ijābiyyati allatī tarakat-hā mubādaratu al-salāmi wa kaḏālika al-<udwānu al-'isra'īliyyu <alā lubnāna bayna ṣufūfi al-yahūdi al-'amirkiyyina..

ṯāliṯan: 'inna jiddiyyata al-mawqifi al-'amīrkiyyi ḥiyāla al-salāmi yumkinu 'ilqā'u al-ḍaw'i <alay-hā 'iḏā mā qāran-nā ḥajma al-musā<adāti al-'amirkiyyati li-miṣra bi-ḥajmi tilka al-musā<adāti allatī qaddamat-hā 'amīrkā li-al-duwali al-'uxrā..

rābi<an: muḥawalatu 'iqnā<i al-'idārati al-'amīrkiyyati bi-'anna-hu 'iḏā kānat turīdu min al-<arabi ḥaqqan qabūla mubādarati rīgin bi-lā ziyādatin 'aw nuqṣānin fa 'inna al-ṭarīqa al-mu'addiya 'ilā ḏālika yajibu 'an yaqūma <alā rasmi <ilāqatin ṯunā'iyyatin mubāširatin bayna al-filastīniyyīna wa 'amīrka..

[I shall take the opportunity of President Mubarak's planned visit to America to consider more closely what I have mentioned earlier in this article and express how I envisage the future steps of the Egyptian diplomacy concerning the Palestinian issue.

First: To curb the antagonistic (anti-Egyptian) propaganda in American society, which attempts to mutilate Egypt's true attitude towards peace; peace which has become an accepted pan-Arab plan, particularly after Fez summit.

Secondly: To capitalise on the positive effect made on the Jewish community in America by the (Egyptian) peace initiatives and the Israeli invasion of Lebanon ...

Thirdly: To shed light on the seriousness of the American attitude towards peace by comparing the size of American aid to Egypt with that offered to other countries ...

Fourthly: To persuade the American administration that the way to convince the Arabs to accept Reagan's initiative is by having bilateral relations with the Palestinians ...]

(Hr, 24/1/83, X27, 5769-813)

In these excerpts the enumerative connectives make a clear demarcation of the various aspects of the argument.

The order of the enumerated (sets of) propositions is sometimes neutral, i.e. the listed items are mutually replaceable. Often, however, the sequence is determined by rhetorical considerations. One such is scaling, where listed items are graded in order of prominence. Additionally, temporal, spatial and logical factors can direct the order to a considerable extent. (Examine, for instance, the ordering of the items in the two excerpts above).

A sequence of propositions connected via enumeration usually has a distinct organisation with an initial stage, i.e. the first proposition or set of propositions, a middle stage (in multi-stage sequences) referring to the next listed proposition(s), and a closing stage, i.e. the final culminating proposition. The initial stage is usually signalled by: English: first, firstly, initially, first of all, for a start, to start (with), to begin (with); Arabic: "bidāyatan" [to start with, first], "'awwalan" [first(ly)], "'awwala mā" [first], "bādi'a ḏī bad'in" [first, to start with, at first], "fī al-bidāyati" [at first, to start with], "fī al-marrati al-'ūlā", [at first]. The middle stage may be signalled in English by: second, third, thirdly, then, and; and in Arabic by "ṯāniyan" [secondly], "ṯāliṯan" [thirdly], "rābi<an" [fourthly], etc., and "wa" [and]. These connectives may also signal the closing stage if it is the final in the enumeration process. Other connectives that specifically signal a closing stage are: in English: last, lastly, finally, ultimately; and in Arabic: "'axīran" [last, lastly, at last], "fī al-xitāmi" [finally, in the end], "fī al-nihāyati" [finally, in the end].

6.4.4.3 Some Textual Patterns

1. In marking the ordering of listed propositions, enumerative connectives may occur in combination, each signalling one stage. In each combination connectives may form a set of two or more different connectives, such as in English: First, secondly, finally (in the excerpt in [6.20] above), First, Then, And in this excerpt:

> [6.22] There is, however, growing disquiet about the way the western alliance has appeared to be pursuing a policy of aggressive confrontation with the Soviet Union. First, the US Government failed to ratify the Salt II treaty signed by Presidents Carter and Brezhnev. Then, it stopped non-aligned initiatives at the United Nations to freeze all nuclear arms development. And it has effectively blocked serious negotiations by intransigently sticking to the so-called zero option ..., which is manifestly unrealistic and one sided.
> (Ti, 6/4/83, X189, 30598-611)

Alternatively, one connective in the combination may be repeated to signal more than one stage, usually the middle and closing ones, as in the following excerpt where "then" is repeated twice to signal the medial and final stages of enumeration.

> [6.23] There are three things wrong with Michael Foot's appearances on television ... First, he ought to get rid of those terrible oxy-acetylene welder's glasses. Then he should stop fidgeting. ... Then he should stop rambling on all the time .
> (O, 24/4/83, X52, 8215-23)

In Arabic the combination is far more restricted. The tendency is to use serial enumeration: "'awwalan" [first], "ṯāniyan" [secondly], "ṯāliṯan" [thirdly], "rābi<an" [fourthly], etc. Refer to the excerpt in [6.21].

2. However, the order of the sequence may not be so explicitly indicated. It suffices in certain cases to start the initial stage

and leave the rest of the propositions implicitly enumerated, or signalled by other structural or lexical means such as the anaphoric expressions "the other" or "the second". For example:

[6.24] So what are the snags? The Treasury discussion paper hits them all. First the loss of revenue
...
The other major objection, says the discussion paper, lies in the fact that companies would be allowed to accumulate financial assets tax free.
(DTel, 13/4/83, X102, 17182-97)

In the following excerpt in Arabic, two reasons are enumerated to justify the view expressed at the beginning, the first one is left implicit while the second is signalled by "taniyan".

[6.25] wa yarā xubarā'u al-maṣrifi 'anna hāḏā al-raqma yumkinu 'i<tibāru-hu 'aqalla min al-wāqi<i li-'asbābin šatta min-hā 'anna al-ma<lūmāti al-xāṣṣata bi-<āmi 1980 gayru mumkinatin, wa ṯāniyan li'anna al musā<adāti al <arabiyyata li-ṣāliḥi 'afriqya lā ta'xuḏu fī al-ḥusbāni al-musāhamāti al-<arabiyyata fi mizāniyyati al-munaẓẓamāti wa al-hay'āti al-duwaliyyati wa allati tufidu -- juz'iyyan -- 'afriqya.
[Financial experts in the Bank believe that this figure can be considered smaller than the real one for various reasons; among them is that sufficient information concerning 1980 is not possible to obtain, and secondly because (the figures indicating the size of) Arab aid to Africa do not take into consideration the Arab contributions to the budgets of international organisations and committees that provide a partial aid to Africa].
(Ar, 26/5/83, X200, 34509-17)

Another fairly common method of signalling enumeration in Arabic is the use of anaphoric items that indicate serial counting, e.g. "(al-sababu) al-'awwalu" [the first (reason ...)], "al-ṯānī" [the second ...], etc.

3. Textually, there seem to be two types of enumeration patterns: discrete and continuous. Discrete enumeration lists the

items systematically, normally through counting: first(ly), second(ly), etc. This type is useful when the text producer would like to compartmentalise the propositions giving each a specific tag and imposing autonomy to its conceptual domain. This is exemplified in the excerpts [6.20-1]. Similarly, the following excerpt from Text (12) in the corpus illustrates this type of organisation. The writer makes a list of the British Treasury's arguments against participating in the EMS exchange rate mechanism, producing a counter-argument as he lists each. Note that since each argument and its counter-argument fill a paragraph, it is necessary to keep them related by repeating "it was argued that".

[6.26] First it was argued that Britain could only join the EMS in the unlikely event of an exchange rate which gave British exporters a crucial competitive advantage over EEC rivals ...

Second it was argued that defence of a fixed rate would produce an unacceptable explosion of the money supply ...

Third, it was argued that British participation in the EMS exchange rate mechanism would be disruptive because of the enormous and unpredictable movements in what is petrocurrency..

(G, 17/11/83, X12, 1682-701)

In Arabic the list of items may be very long. Counting often run to four (a pattern not discerned in the English corpus), occasionally to six or even seven, and in one case counting goes up to eleven (Text 9 of the Arabic corpus which contains a discussion of George Shultz's "blunders" in handling the Middle East crisis during 1982-3).

To make enumeration tighter, the sequence is often introduced with a statement specifying the number of items to be enumerated. The excerpt in [6.23] gives an example (in English) where enumeration starts by "There are three things wrong with Michael

Foot's appearances ...". Then each of these items is clearly marked with an appropriate enumerative. Such a pattern is, however, less often in Arabic, the tendency being to make an introductory proposition but only occasionally specify how many items are to be listed. The following excerpt is one of those infrequent cases:

[6.27] al-'ijrā'ātu al-siyāsiyyatu 'inṣabbat bi-šaklin markaziyyin <alā taḥmīli al-wilāyāti al-muttaḥidati al-'amīrkiyyati mas'ūliyyata mā ḥadaṯa fī bayrūta li-sababayni 'iṯnayni:
'awwalan: bi-sababi al-<ilāqāti al-xāṣṣati allatī tarbuṭu al-wilāyāti al-muttaḥidata bi-al-kiyāni al-ṣahyūniyyi.
tāniyan: li'anna al-wilāyāti al-muttaḥidata lam tafi bi-al-ḍamānāti al-maktūbati al-muqaddamati li-al-muqāwamati al-filasṭīniyyati wa al-muta<alliqati bi-<adami al-samāḥi li-al-ṣahāyinati bi-al-duxūli 'ilā al-muxayyamāti al-filastīniyati.
[The political measures adopted have concentrated directly on relating the responsibility of what happened in Beirut to America, for two reasons:
- First: because of the special relations that tie the United States with the Zionist entity (Israel).
- Second: because America failed to meet its assurances that it will not allow the Israelis to enter the Palestinian camps ...]
(Th, 13/10/82, X65, 11819-33)

The continuous type of enumeration, as opposed to the discrete one, is not so directly itemised. The normal pattern utilises the enumerative force of the connective to indicate one stage, usually the initial, though occasionally the final. The pattern starts with a proposition introduced by a connective such as first, first of all, for a start, to start with (in English) or "bidāyatan", "fī al-bidāyati", "'awwalan" [first, first of all] (in Arabic). Once the first stages in the enumeration of arguments is established, the rest of the stages are allowed to run without necessary itemisation. The following excerpts exemplify the first stage. (The text can be consulted for tracing the rest of the arguments).

[6.28] But the future of the Labour Party interests me today less than that of the Alliance, and especially the SDP half of it. To start with, Mr Steel is going to need all his skill at diplomacy to prevent the more idiotic of his followers (such as the ones in the Liverpool Broadgreen constituency, for instance, who insisted on putting up a Liberal against the SDP candidate agreed by the leadership of both parties) from insisting that the Alliance should be broken up immediately ...
(Ti, 15/6/83, X219, 34402-11)

[6.29] bidāyatan yaṣiḥḥu al-qawlu bi-'anna-hu lawla al-waḍ<u al-<arabiyyu al-mutaraddi la-mā 'asrafa al-kiyānu al-ṣahyūniyyu fi 'i<tidā'āti-hi wa jarā'imi-hi al-barbariyyati, wa la-mā 'aqdama <alā gazwi al-'arāḍi al-lubnāniyyati, wa la-mā 'istaṭā<a 'ilḥāqa al-'aḍā bi-al-muqāwamati al-filasṭīniyyati wa ḍabḥi al-'ālāfi min 'abnā'i-hā.. [First, it is only correct to state that had it not been to the general Arab deteriorating (and inconsistent) policies,the Zionist entity (Israel) would not have dared to continue its barbarian crimes, nor would it have invaded the Lebanon or inflicted severe damage to the Palestinian resistance movement and slaughtered thousands of the Palestinians ...]
(Th, 13/10/82, X65, 11862-8)

Although this type of enumeration has been labelled "continuous", it is occasionally disrupted by the introduction of various other textual patterns that build up the rest of the text, for instance the expression of cause and effect, discrepancies, and various textual orientative moves. Hence this pattern may suffer from looseness in item listing.

6.4.5 Amplification

6.4.5.1 Repertoire

The following list comprises connectives identified in the corpus as signalling an additive relation of amplification.

A. English

Simple: and, even,

Compound: above all, most of all

Correlate: not only ... but also

B. Arabic

Simple: bal, ḥattā, 'innamā, wa

Compound: wa law

Correlate: (lā, lam, laysa)..bal, (lā, lam, laysa)...lākin, (lā, lam, laysa)...lākinna

6.4.5.2 Textual functioning

The additive relation of amplification is basically enumerative in the sense that two or more related blocks of knowledge are listed. The difference, however, lies in the fact that the last proposition represents an intense degree or phase in relation to the previous ones(s). The main function of the connective is to alert the audience that the forthcoming proposition is by far the most prominent in the series enumerated. This function has an important role in text organisation: it shapes the order of the arguments, stating the comparable ones first and delaying the prominent one to the end where, with the help of the connective, it is placed under focus. Examine these examples in English and Arabic.

[6.30] So she softened her hairstyle, lowered her voice and chatted on endlessly about minced beef and her favourite colour. She submitted herself to such journalistic indignities as the interview in the Sun under the heading: "May face, My Figure, My Diet" where we learnt about he issue of the Thatcher Double Chin. Above all, she fostered the image of the woman as good house-keeper who would look after the nation more efficiently than any man.

(G, 8/6/83, X42, 6653-6662)

[6.31] There is nothing flamboyant about "Bernie" Rogers. He does not carry a pearl-handled revolver on his hip. He does not even look like a general ...

(DTel, 22/2/83, X97, 16443-45)

[6.32] ..hāḏihi al-naẓratu al-sīkūlūjiyyatu tu'aṯṯiru <alā ḥarakati al-mu<āraḍati, bal tuhaddidu-hā bi-<adami al-najāḥi.
[This psychological view affects the opposition movement, even threatens it with failure]
(Ar, 17/6/83, X212, 36568-70)

6.4.5.3 Textual Patterns

There are two notable patterns of connectives involved in signalling this relation: one is strong and depends on the semantic force of the connective itself, and the other is weak and relies on the semantic environment of the propositions. In the first, the propositions are listed and the connective explicitly places the prominent one under focus (see examples 6.30-32) above). The connective can be so forceful at times that the way it stresses the prominent proposition is intensified to the extent that the previous propositions in the list are greatly subdued, resulting in creating a shift in meaning that resembles that of an antithetic semantic relation. This is particularly true with the use of the connective "even", which is why grammarians such as Quirk et al. (1972, 1985) have identified two aspects of its meaning: additive and adversative. It is also true of the connective "bal" in Arabic. The following excerpts are illustrative examples.

[6.33] I [would] organise my life to live in Amiens if you gave me an English university library and a chair of my own to sit on. I am not even asking for an English tutor's bottle of sherry ...
(G, 14/12/82, X7, 990-3)

[6.34] ..'inna al-'ittifāqa al-lubnāniyya al-'isrā'iliyya laysa 'aḵṯara min faxxin 'isrā'iliyyin 'amīrkiyyin li-tawrīṭi lubnāna fī ṣirā<atin dāxiliyyatin wa <arabiyyatin wa tafjīri-hi bi-sababi al-tawqi<i <alā 'ittifāqin yulzimu-hu wa lā yulzimu 'isrā'īla, wa yaj<alu al-tafāwuda bayna bayrūta wa dimašqa, 'iḏā ḥaṣala, yatimmu wafqa al-šurūṭi al-'isrā'iliyyati. bal yaj<alu al-wifāqa al-lubnāniyya al-lubnāniyya yatimmu 'ayḍan wafqa hādihi al-šurūṭi..

[The Lebanese-Israeli agreement is no more than an American-Israeli trap to drag Lebanon into internal and external Arab conflicts caused by signing an agreement that commits Lebanon but not Israel and that causes and negotiations between Beirut and Damascus, if they are to take place, to be brought about according to the Israeli conditions. It even causes the Lebanese-Lebanese concord to be brought about in accordance with these conditions]
(Nb, 22/5/83, X15, 2513-21)

In the second pattern, the weak one, the textual environment, particularly the sequencing of the items, highlights the function of amplification. Here the way the propositions are arranged in the list pushes the most prominent one to the end of the list, appending it additively to the previous one via the connective "and" in English or "wa" in Arabic. If you introduce a change in the environment by shuffling the ordering of the propositions, the connective "and" loses its amplificatory function and remains an additive by appending. The following excerpts are examples from the corpora:

[6.35] We don't want our girls to be the victims. We want to help them recognize the unremitting pressure to accept value-free sex; to believe that there are some things worth waiting for. And to understand that sex is too good to be spoilt by casual affairs.
(DTel, 13/4/83, X103, 17338-17343)

[6.36] ..kaṯiran mā nasma<u bi-'an yuqāla bi-'anna al-mudīra al-<āmma al-fulāniyya 'aw ra'īsa al-mu'assasati al-fulāniyyati rūtīniyyun 'aw šadīdun jiddan, wa rubbamā yuqālu "zālimun" 'aḥyānan..
[We often hear it said that a director general is rigid or strict, and many (even) say "unfair"]
(Th, 5/4/83, X73, 14067-14070)

6.4.6 Comment

6.4.6.1 Repertoire

The following short list comprises connectives that express the additive function of comment.

A. English

Simple: and, which

Compound: after all

B. Arabic

Simple: wa, fa, mimmā, ḥayṯu.

Compound: 'alā wa.

6.4.6.2 Textual Functioning

Despite the short number of connectives that signal the additive function of comment, their role is vital in organising stretches of text additively. The label "comment" should be taken in a broad sense to subsume not one specific function, particularly in Arabic, but rather, a group of related functions that collectively have one general role. It gives the text producer the chance to pause and comment or express a view regarding the previous proposition. The functions covered by "comment" are:

1. It introduces a definition or explanatory statement or an explanation of either the whole content of the previous proposition or one of the concepts included in the proposition.

[6.37] Some foreign diplomats are reported to wonder about Mr. Reagan's ear for the nuance and detail of technical issues and the arguments of adversaries, which is a polite way of saying that they fear he'll muck it up.
(O, 29/5/83, X73, 11948-52)

[6.38] ..lā budda la-hu min wujūdi al-maṣdari al-fikriyyi alladī yūḥī 'ilay-hi bi-al-<amali al-fanniyyi. wa hādā al-maṣdaru huwa al-munāsabatu 'aw al-ḥadaṯu 'aw al-mawqifu 'aw al-manẓaru..
[There must be a mental source that inspires him with artistic works. And this source is the event, the context or the scenery.]
(Hr, 25/1/83, X30, 6079-82)

2. It introduces a commentary on the last proposition. This commentary, in most cases, expresses the text producer's stand in relation to the statement made, or it may announce a judgement, either personal or general, of the content of the statement. The following are illustrative examples:

[6.39] The only motive for reading it may be to swank about it at literary cocktail parties, which is a wet and wimpish reason.
(Ti, 28/5/83,X208,33081-33084)

[6.40] ..al-'ijābatu bi-al-ta'kīdi na<am. wa hiya 'asbābun mawḍū<iyyatun wa wāḍiḥatun tasabbabat fī al-balbalati al-fikriyyati..
[The answer is definitely yes. And these are objective and clear reasons that have caused intellectual confusion.]
(Nb, 14/7/83, X20, 3649-3651)

3. A function, related to the previous one, expands the judgement to an explanation based upon, or an inference deducted from, the current propositions. Although there is an element of causality, the relation is basically additive. Examine the use of and and "mimmā" in these two excerpts respectively.

[6.41] For a long time we were unable to recognise the nature of this violent disturbance of our lives, for reasons that are obvious. It promised us release from an earlier, scarcely bearable poverty; and this is why our resistance was lowered.
(G, 17/1/83, X13, 2029-2033)

[6.42] laqad qarrara al-<arabu al-bahta <an al-salami fi al-<awasimi mimma yu'akkidu salafan al-taxalli <an sigati al-nidali al-taharruriyyi.
[The Arabs have decided to look for peace in the capitals, which confirms in advance their abandoning of the policy of liberation struggle.]
(Nb, 22/5/83, X15, 2323-2325)

4. Another function is parenthetical. The connective introduces a statement that interrupts the flow of propositions to present a

proposition, deemed by the text producer as essential, but may not be so to the development of the textual sequence. Because of their interruptive nature, such propositions are often separated orthographically from the other sentences with dashes.

[6.43] If you add the Alliance's seven and three quarter million votes to Labour's 119 lost deposits, you will get only one possible answer - and I speak as a man whose brother-in-law is the Professor of Mathematics in Amherst, Massachusetts.
(Ti, 15/6/83, X219, 34445-49)

[6.44] wa ya<taqidu al-ba<ḍu -- wa al-<abdu li-allāhi min-hum -- 'anna al-ḥakīma ma zāla wā<iyan jayyidan li-'aqwāli-hi..
[And some believe - and I am one of them - that Al-Hakim is still fully aware of his views.]
(Nb, 4/4/83, X3, 218-20)

The four variant functions of the additive relation of comment exist in both languages. However, there are two other variants that are distinctly peculiar to the Arabic language. These play an essential role in organising propositions into bigger sequences.

5. The connective has the function of signalling the main viewpoint or block of information. It is a function of alerting the audience (as well as the processor) that the next proposition is the focus of attention in relation to the current stretch of text. Connectives that perform this function are "fa" and "wa", which exhibit in their use certain syntactic constraints. The connective "fa" in particular signals this function when it occurs after adverbial modification (e.g. temporal, causal or adversative), after some connectives and after certain subordinate clauses. The clause that "wa" or "fa" introduces is usually nominal (the first element is a substantive or the use of the particle "'inna").

[6.45] wa mā <adā-hā fa lam takun 'akṯara min "'aḥkāmin" qaḍay-tu-hā bayna ḥurūbin wa hudnātin wa hurūbin.
[And with the exception of that, it was not more than "sentences of imprisonment" that I spent between wars, truces and (yet more) wars].
(Nb, 4/4/83, X2, 101-3)

[6.46] mundu fatratin wa al-'isra'īlīyyūna yudīr-ūna ḥamlatan 'i<lāmīyyatan <an al-ṣawārīxi al-sūfītiyyati.
[For some time now, the Israelis have been conducting a campaign against the Soviet missiles (in Syria)]
(Sh, 20/4/83. X124, 24077-78)

6. The additive "wa" may occur immediately after certain connectives in order to intensify their textual function. "Wa" in this position maximises the meaning and textual role of the preceding connective and places the next proposition in focus. Connectives that have been observed intensified in this way in the corpus are: "bal" [but, even], "xāṣṣatan", "xuṣūṣan", "siyyamā", "bi-al-xuṣūṣi" [in particular, especially], "'illā" (when used as a correlate connective).

[6.47] ..'inna al-ṭifla sawfa yaḥtaqiru-nā. bal wa qad yaṣilu 'ilā 'iḥtiqāri nafsi-hi.
[... the child will despise us. He may even start to despise himself ...]
(J, 5/2/83, X84, 16632-34)

A related function that "wa" achieves (which is, incidentally, not very common in the corpus) involves specification of the function of a connective. This is done by chaining the connective to adverbial phrases or other connectives in order to delimit its perspective and delineate its scope of operation.

[6.48] 'iḏan wa bi-lā kaṯrati šarḥin wa taṭwīlin, ..
[Therefore, and without too much (unnecessary) explanation or expansion ...]
(Sh, 20/4/83, X124, 24062)

6.4.7 Continuity

6.4.7.1 Repertoire

A. English

Only one connective has been identified in the English corpus to signal this function: and.

B. Arabic

In Arabic, the list comprises the following:

Simple: wa, 'aqūlu, fa, kamā, naqūlu

Compound: hāḏā wa, wa ba<ḍu

6.4.7.2 Textual Functioning

Although text continuity is sustained through the use of all types of connectives, the additive connectives of continuity define a particular role that is vital in organising text in Arabic. That does not mean that additive connectives of continuity do not exist in English; but due to their limited use and the fact that their text-forming function is conflated with main appending functions, one can safely say there are no additive continuative connectives that create text in English in the same way they do in Arabic.

Additive connectives of continuity are operational when they link propositions or sequences of propositions that, though related in the text world knowledge, are independent enough to discard with any form of concatenation via connectives. A typical case is when a knowledge configuration that is expressed in a proposition or set of propositions and that is self-contained is terminated and a new configuration is attempted. In Arabic, the move between the two sets will at least require an additive connective to signal continuity; in English continuity is assumed as default. A common

case is the boundaries between paragraphs. A new paragraph in Arabic requires an additive-continuative to introduce it. For example,

[6.49] ..wa lākinna jūrj šūltz kamā 'aṯbatat al-'aḥdāṯu ṭilata al-'asābi<i al-'arba<aṯi al-māḍiyati huwa 'agbā wazīri xārijiyyatin fi tārixi al-wilāyāti al-muttaḥidati al-'amīrkiyyati <alā al-'iṭlāqi!
wa qad jā'a jūrj šūltz 'ilā bayrūta muta'axxiran <an mi<ādi-hi al-munāsibi 'akṯara min xamsati šuhūrin!
[... but George Shultz, as the events of the last four weeks have proved, is definitely the most stupid foreign minister in the history of the United States of America.
(And) George Shultz arrived in Beirut five months later than he should appropriately have done ...]
(Nb, 20/5/83, X9, 1013-20)

This excerpt represents the final and first parts of the first and second paragraphs of a long article on George Shultz's policies in the Middle East. Note that the second paragraph is introduced with the additive connective "wa" to impose a sense of continuity between the two paragraphs.

Related to the additive-continuative function is one in which the connective "wa" intensifies continuity by providing support to other connectives. This takes place when "wa" occurs immediately before a connective and serves to maximise the smoothness of the continuity of the concatenation. In this function, "wa" can be associated with most Arabic connectives. In the corpus, a large number of connectives are supported by "wa", e.g. "lākin", "lākinna", "innamā", "hākaḏā", "bi-al-tālī", "<alay-hi", "lawlā", "li-hāḏā", "ka'anna", "ma<a ḏālika", "law", "mahmā", "'axīran", "ḥattā", etc. (See also Chapter 9 and the relevant discussion and table). Examine this excerpt:

[6.50] wa al-'amṯilaṯu min al-māḍi al-qarībi wa al-ba<īdi kaṯiraṯun kaṯiraṯun; wa lākin-nā naktafī bi-al-miṯālayni al-'axīrayni.
[And the examples (that can be) drawn from the near and distant past are numerous; but we will suffice ourselves with the last two ones.]
(Sh, 18/3/83, X119, 23001-3)

where "wa" gives considerable support of continuity to the adversative connective "lākinna" [but]. In English, such support is restricted:

[6.51] This popular attitude is nearer to Mr Denis Healey's position than to Mrs Margaret Thatcher's. And yet the majority, according to the polls, greatly prefer the Conservatives' defence policy to Labour's.
(O, 29/5/83, X78, 12899-903)

6.4.7.3 Textual Patterning

In Arabic the additive connective of continuity "'aqulu" (I say) is used when the accumulation of propositions is felt (by the text producer) to have blurred the main core or central proposition. The writer then pauses, uses the connective to resume what he has started to say.

[6.52] laysa ṣaḥīḥan 'anna wujūda quwwatin mutamayyizatin bi hawiyyati-hā wa šaxṣiyyati-hā al-ḥaḍāriyyati qad yuxriju al-ḥiwāra min jawwi-hi al-hādi'i al raṣīni wa qad yurdī bi-wujūdi tilka al-quwwati nafsi-hā li'anna... 'aqūlu laysa hāḏā ṣaḥīḥan li'anna...
[It is not correct that the existence of a power distinguished for its cultural identity would ... I say it is not correct because ...]
(Sh, 16/6/83, X140, 27091-99)

where a relatively long stretch of text intervening in the middle makes it necessary to pause and use the connective to introduce adjustment that ensures continuity and thus be able to resume the earlier proposition that the whole sequence has started with.

The connective "wa ba<du" is used when a gap is felt to have

occurred in the arrangement of the knowledge blocks. The connective is used to bridge the gap to ensure continuity of two distantly related knowledge configurations. Because of the nature of this function, some writers prefer to use it in a paragraph of its own to mark clearly the process of adding the previous stretch of text to the next one. Examine this example:

[6.53] wa-ba<du
fa tilka hiya qaḍiyyatu al-ḥurriyyati -- 'arḍan wa 'insānan..
[Then.
That is the cause of freedom - land and man ...]
(Sh, 26/5/83, X130, 25590-93)

Note that this connective is usually followed by "fa" used as an additive connective of comment. Note also that in the corpus the connective is used towards the end of a text bridging the main arguments with a conclusion.

The most frequent connective of continuity "wa" often bridges the gaps between paragraphs to ensure continuity. Hence, with the exclusion of the first paragraph in a text, few paragraphs are not introduced by this connective. Alternatively, the connective "wa" may bridge two large sequences of propositions within a paragraph, or two distinctly independent propositions.

6.4.8 Coupling

6.4.8.1 Repertoire

This category has been observed to operate in Arabic. Connectives that signal this function are "wa" and "fa", though "'aw", which otherwise signals alternative relations, has also been identified.

6.4.8.2 Textual Functioning

This additive function is basically of the appending kind (see

6.4.3). However, it differs in the type of propositions combined and the overall rhetorical effect required. We have observed in the corpora that this function is predominantly peculiar to Arabic (cf. Al-Jubouri 1984).

The connected propositions have normally similar or at least closely related semantic content. The concatenation is made mainly for rhetorical effect: to create forceful assertions via repetition of structures and propositions. For instance, in the following excerpt

[6.54] ṯumma rafa<a al-'ixwatu al-sūriyyūna nafsa al-ši<āri wa bada'-ū yumāris-ūna-hu wa yuṭabbiq-ūna-hu kamā yatarā'a la-hum!
[Then our Syrian brothers raised the same slogan and started to apply it the way they liked!]
(Nb, 14/7/83, X20, 3753-55)

the verbs "yuṭabbiq-ūna" and "yumāris-ūna" share a similar meaning: "to apply" (cf. Wehr's Dictionary). Similarly in

[6.55] ..'anna 'amirka qad 'a<lanat mawqifa-hā wa qālat kalimata-hā wa al-bāqī matrukun li-al-<arabi!
[... that America has announced its position and made its statement ...]
(NB, 15/7/83, X22, 4485-6)

[6.56] ..ḥattā sakata al-nabḍu wa xamadat al-'anfāsu.
[... until the pulse stopped and the breath ceased (=until he died).]
(Hr, 25/1/83, X32, 6371-2)

the two clauses in each example share the same contextual meaning within the text.

Further, in some cases the repetition has lost its rhetorical effect because of frequent use:

[6.57] 'a<ūdu fa 'aqūlu..
[I repeat ...]
(J, 5/2/83, X82, 16167)
(Ar, 30/6/83, X220, 37402)

The use of "aw" [or] intensifies this function by involving the

two propositions in a pseudo-alternative relation. For example, in:

[6.58] ..mā yumkinu 'an yusfira <an-hu 'aw 'an yu'addiya 'ilay-hi min natā'ija..
[... the consequences it could lead to ...]
(Ar, 21/1/83, X180, 31418-9)

the verbs "yusfira" and "yu'addiya" share the same contextual relation to "natā'ija" [consequences]; both of them imply "lead to".

6.5 Comparison

6.5.1 General Comments on Textual Functioning

This relation obtains when one knowledge configuration is delimited by being measured against and compared to another. Generally, there are two distinct types of comparison. The first corresponds to the mathematical symbols > and <, the second to = or ≡. In the first type the current proposition has a conceptual centre that is comparatively big or small compared to the corresponding conceptual centre in the following proposition. In the second type, the two propositions exhibit identicality or similarity of conceptual configuration. We have to allow for a third type where the comparison is indeterminate or at least not as clear-cut as it is in the other two types.

Comparison is achieved with the help of a relatively small number of connectives that not only indicate the type and scope of the relation but point to the direction of textual emphasis. Normally, emphasis is achieved in a progressive forward direction, in which the current proposition (the antecedent, the one that is not introduced by the connective) is highlighted and made prominent by being conceptually contrasted to the following proposition (the subsequent, introduced by the connective). But occasionally,

textual emphasis is regressive and looks backward; that is, the consequent is highlighted when a central conceptual configuration is made explicit by being defined in terms of the previous proposition. This function, as an examination of the two corpora has suggested, seems to be peculiar to English. The following are illustrative examples. (The first two exemplify the progressive function of comparison in English and Arabic; the third excerpt exemplifies the regressive function in English.)

[6.59] Our problems are very deep-rooted, as indeed were France's before the war.
(DTel, 29/1/83, X96, 16372-4)

[6.60] fa hal naḥnu bi-lā mustaqbalin mitlamā naḥnu bi-lā ḥāḍirin?
[Are we without any future as we are without any present?]
(Nb, 4/4/83, X1, 81-2)

[6.61] As for the moral lead implied by unilateral disarmament, this, too, would call for inspirational leadership of truly Periclean proportions. Likewise on the home front, Labour's programme for curing mass unemployment envisages a quality of public intervention and co-ordination the like of which has not been seen since 1940.
(STel, 22/5/83, X131, 21910-16)

The function of comparison can be subcategorised into two closely related functions, depending on its scope and the textual motivation behind its use:

a) comparison of similarity, where (i) the scope extends and includes all the conceptual configuration in the two propositions and (ii) interest lies in establishing similarity (or otherwise) without a conscious effort to measure its extent or intensity.

b) comparison of degree, where (i) the scope of comparison is restricted to one aspect in the knowledge configuration of the two propositions, and (ii) the main purpose is to establish the degree

or extent of similarity (or dissimilarity).

Both types are discussed below.

6.5.2 Similarity

6.5.2.1 Repertoire

The following connectives signal a relation of similarity:

A. English

Simple: as, equally, likewise, similarly

Compound: as if, as though, in similar fashion, in a similar manner, in a similar way, in a way (that), in much the same way as, (in) the same way, in the same way as, in this manner.

Correlate: (just) as ... so, such ... as

B. Arabic

Simple: kamā, ka'anna, miṯlamā, kaḏalika, ka'an, ka'annamā

Compound: bi al-miṯli, bi-miṯli mā, <alā hāḏā al-minwāli, <alā hāḏā al-wajhi

Correlate: kama...fa

6.5.2.2 Textual Functioning

The main function of this category is to signal a similarity (or otherwise) in the conceptual configuration of two propositions, thus bringing them closely together. On closer examination of this function as signalled by the connectives, one can observe a multiplicity of sub-functions:

a) The core function is comparison proper, where the text producer establishes two equal, identical or similar conceptual configurations in the textual world. The examples in [6.59-61]

above illustrate this function.

b) A related function signals similarity of the manner in which two events or courses of action take place. In other words, it indicates that an event or course of action in one proposition took, or will take, place in a similar way to another event in another proposition. For example

[6.62] ... the police presence was meant to symbolise the State's determination to play a leading part in this visit and not to retreat off-stage as it did in 1979.
(O, 19/6/83, X83, 14021-4)

[6.63] ..'amīrkā sa-tamḍī fī musā<adati lubnāna kamā tusā<idu miṣra wa 'isrā'īla..
[America will carry on giving aid to the Lebanon as it gives aid to Egypt and Israel.]
(Nb, 15/7/83, X22, 4482-3)

c) In displaying similarity, the consequent may give exemplification or evidence for the content of the subsequent proposition. For example,

[6.64] ... in some power struggles the Central Committee has played a vital role - as when Khrushchev ousted his rivals, Melenkov and Molotov, the so-called "anti-party group", or in Khrushchev's own downfall.
(Ti, 13/11/82, X160, 26279-83)

[6.65] ..lan yusajjil-ū li'anna-hum lā yaḍman-ūna ṣarfa al-mustaḥaqqāti miṯlamā ḥadaṯa fī 'ugniyatin sābiqatin.
[they (the music band) will not make a recording because they cannot guarantee receiving their fees as it once happened with one of their previous songs.]
(Hr, 24/4/83, X52, 9571-3)

d) Similarity can be hypothetical. The subsequent proposition is compared to a hypothetical one for illustrative purposes or to create more textual salience for the content of the subsequent proposition. This is an old rhetorical device and has its roots in

rhetorical imagery. A hypothetical proposition should be understood as one that does not exist in the textual world or perhaps runs contrary to world knowledge. The examination of the two corpora has shown that this is one of the main functions of similarity, particularly in Arabic. Connectives that signal this function in English are: as, as if, and as though. In Arabic, it is signalled exclusively via the connectives "ka'an", "ka'anna", "ka'annamā" and "kamā law", the first three are usually preceded by a supportive additive-continuative "wa". The function is also occasionally signalled by "kamā" and "miṯlamā". Examples:

[6.66] The egg develops as if it had been fertilised by a sperm cell ...
(G, 2/12/83, X6, 798-9)

[6.67] ..fa qad warada fī al-kalimati al-xitāmiyyati li-al-taḥqīqi 'anna 'iḥdā al-jam<iyyāti tursilu al-musā<adāti li-al-xāriji "wa ka'anna mašakila-nā qad 'intahat".
[It was mentioned in the final part of the report that one of the charity funds sends aid abroad as if all our problems are non-existent.]
(Nb, 5/4/83, X6, 722-25)

e) Similarity may sometimes be conflated with additivity. In such cases the predominant function is still the establishment of comparison but the additive meaning is obvious. In English this function may be signalled by the connectives similarly, equally, likewise, and less frequently, as. In Arabic, it is signalled by the connectives "miṯlamā" and "kamā".

[6.68] The liberal conscience is never going to be persuaded to accept the death penalty as form of punishment. Presented in that guise, it will always be dismissed as primitive revenge. Likewise, it will always be rejected as a unique deterrent.
(STel, 26/6/83, X148, 24805-10)

[6.69] wa al-'istihānatu bi-hā yušīru 'ilā tadannin fī

mustawā al-'īmāni 'aw al-wa<yi 'aw kilay-hima ma<an miṭlamā tumaṭṭilu 'inxifāḍan fī mustawā tajsīdi al-muwāṭanati al-ḥaqqati..
[...and the lack of respect for them (regulations and new traditions) points to a low level of faith or awareness or both as it represents a low level of the feeling of true citizenship.]
(Th, 5/4/83, X73, 14063-66)

6.5.3 Degree

6.5.3.1 Repertoire

The following connectives express the function of degree:

A. English

Simple: than

Compound: on this scale, to a point where, to a considerable (great, small) extent, to (some, this, that) extent, to an extraordinary (inordinate) degree, up to a point.

Correlative: as ... as, so ... as

B. Arabic

Simple: mimmā

Compound: bi-qadri mā, 'ilā al-ḥaddi, bi-al-qadri, <alā qadri

6.5.3.2 Textual functioning

Comparison by degree establishes a relation whereby a standard is set up, measurable in terms of intensity or degree. Normally one concept acts as the control centre for the comparison and the propositional content of the two clauses are related to it.

There are three types of degree in English that are signalled via connectives, but only two in Arabic. These are: a) equality, which can be symbolised as = or ≡; b) inequality, symbolised as > or <;

and c) indeterminate, where the degree is expressed in general, rather than specific, terms. These are briefly discussed next.

1) Equality: This is signalled by the correlates as ... as and so ... as in English where, structurally, the word representing the concept to be compared is inserted in-between. This helps the processor to establish a short-cut path for setting up the relation and accessing the next proposition.

[6.70] The nuclear deterrence issue seems to be fading from this election as fast as it appeared ...
(G, 7/6/83, X40, 6301-2)

In the Arabic corpus, we have not been able to identify a connective that operates in the same way. The relation is realised via other structural units, usually prepositional phrases. However, the Arabic connective "bi-qadri mā" [as much as] can establish a kind of equality of degree. For example,

[6.71] 'inna-hum yataḥāyal-ūna <alā wāqi<i-him al-jugrāfiyyi bi-al-qawli bi-'anna-hum hum nihāyatu al-šarqi bi-qadri-mā hum bidāyatu al-garbi.
[They (the Japanese) make use of their geographical position by claiming that they are the end of the East as much as they are the beginning of the West.]
(Nb, 4/4/83, X2, 125-8)

2. Inequality: Although inequality can be signalled by negating the equality as established by as ... as, so ... so, it refers here mainly to the relation signalled by than (English) and "mimmā" (Arabic). The two connectives share similar structural patterning and textual functioning. The control centre is represented by a comparative degree adjective in both English and Arabic (though the morphological pattern of the comparative degree is distinctly different). The textual function indicates a variation in degree in relation to the control centre in the

antecedent proposition compared to the subsequent. It should be mentioned here and will be elaborated in the next chapter that there is a substantial difference in the distribution of this pattern in both languages. The following two examples are typical illustration from the English and Arabic corpora.

[6.72] Present welfare arrangements serve bureaucracies better than they serve recipients ...
(DTel, 8/6/83, X121, 20214-16)

[6.73] 'inna al-filasṭīniyyīna laysa laday-him mā yaxsar-ūna-hu 'akṯara mimmā xasir-ū.
[The Palestinians have got nothing to lose more than they have already lost.]
(Ar, 20/4/83, X196, 33746-7)

3. Indeterminateness of Degree: In this type of relation the degree of comparison is not as clearly and specifically determined as they are in the previous two (i.e. equality and inequality). The function here involves setting up of a certain degree or limit as a standard (usually within the content of the antecedent proposition) against which the content of the subsequent proposition is to be viewed, assessed and measured. Therefore there is no direct comparison in terms of degree or intensity. For example, in this excerpt:

[6.74] The consumer has no vested interest in the survival of OPEC as such. But the organisation ... does remain the only means at this time of keeping some stability in oil price movements. To that extent the consumer should wish OPEC well this weekend.
(O, 23/1/83, X48, 7675-80)

the last proposition is meant to be measured against the proposition before it; and this function is dictated by the phrasal connective "to that extent", without which the relation cannot stand. In Arabic, the examination of the corpus suggests that this textual

function is not made via connectives of this type, but rather through prepositional phrases and embedding.

6.6 Alternative

6.6.1 Repertoire

Connectives that signal a relation of alternation, i.e. choice between alternatives, comprise the following list:

A. English

Simple: or

Correlate: either ... or

B. Arabic:

Simple: 'aw, 'am, sawā'an

Correlate: 'immā...(wa 'immā, 'aw)

6.6.2 Textual Functioning

The alternative relation refers to two (or more) facts, events, or courses of actions, both (all) are possible in the textual world but only one is valid or accessible. Logically, the alternative relation (as discussed in Ch. 4 above), is usually of the "exclusive" type, in the sense that at least and at most only one option is true in a possible textual world. This means a) that the textual world in which the options are satisfied must not be epistemically accessible (cf. Van Dijk 1977a), and b) that when one option is valid in a particular textual world, the other must be equally valid in an alternative textual world. For instance,

[6.75] Mr. Sproat must be very badly advised ... or he has deliberately chosen to ignore the actual position...
(DTel, 20/4/83, X107, 17903-5)

[6.76] ..Bīgin yurīdu 'an yusābiqa al-zamana, 'aw yusābiqa al-mawta..
[... Begin wants to have a race with time or have a race with death.]
(Ar, 29/11/82, X161, 29712-3)

This function is made more forceful through the use of the connective "either ... or" or its Arabic counterpart "'immā ... wa 'immā".

The alternative function can be classified into two types according to its place in the textual world and to the semantic nature of the alternatives themselves. Looked at from the first angle, the alternative relation can be of two types: necessary or accidental. A necessary alternative relation is based on conceptual or logical inconsistencies. For example, in

[6.77] The parliament will be presented with a draft which it can either accept or reject.
(G, 28/3/83, X15, 2398-9)

the two propositions "it can accept", "it can reject" are mutually exclusive and hence the setting up of an alternative relation is deemed necessary to establish a correct textual structure. Similarly in Arabic, this cynical excerpt about the Arabs' political state of affairs establishes a necessary alternative relation.

[6.78] yabdū 'anna-nā naḥnu al-<araba 'agbiyā'u jiddan. 'aw rubbamā nakūnu fī muntahā al-ḏakā'i..
[It seems that either we the Arabs are very stupid. Or perhaps we are extremely clever.]
(Ar, 22/11/82, X158, 29475-6)

An accidental alternative relation is based on compatible facts, events or intentions during a specific period of time. For example, in this excerpt:

[6.79] The theory is wrong or it has been poorly implemented.
(Ti, 13/6/83, X216, 33942-3)

the two alternatives are compatible with respect to the text world; i.e. there are no contradictory properties that logically excludes one from the other. However, in the relevant context and period of time only one is valid. The following excerpt is a similar example from the Arabic corpus:

[6.80] al-lubnāniyyu yurīdu 'an ya<rifa matā yastaṭī<u al-<awdata 'ilā qaryati-hi li-yaflaḥa 'arḍa-hu 'aw li-yuqallima 'aš̌jāra-hu al-muṯmirata wa min gayri 'an ya<tariḍa-hu jundiyyun 'isrā'īliyyun..
[The Lebanese citizen would like to know when he could return to his village to cultivate his land or prune his fruit trees without being harassed by an Israeli soldier.]
(Sh, 20/4/83, X124, 23882-5)

6.6.3 Some Textual Patterns

According to the interaction of the semantic content of the two alternative propositions, a number of patterns of the alternative relation can be identified:

1. The two alternative propositions represent independent knowledge configurations. This is the core pattern of the alternative relation. Examine these excerpts:

[6.81] The choice which the economic research is giving Mrs Thatcher is a simple one - she can keep her inflationary gains by having such a weak recovery that unemployment goes on rising, or she can have a real recovery and lose her low inflation rate.
(G, 2/12/82, X5, 723-8)

[6.82] 'aglabu al-ẓanni 'anna-nā nuwājihu mawqifan: 'immā 'an tuḥalla fī-hi al-muš̌kilatu bi-ṣidāmin musallaḥin tatajāwazu xuṭūrata-hu ḥudūda al-<aqli, wa 'imma 'an yatagallaba ṣawtu al-<aqli wa tūḍa<u al-muš̌kilatu 'amāma mu'tamarin duwaliyyin..
[It is most likely that we are facing a situation in which either the conflict is solved via a military encounter whose dangerous consequences go beyond the imagination, or the voice of wisdom prevails and the conflict is placed before an international conference ...]
(Nb, 7/7/83, X16, 3039-43)

In each, the alternative propositions stand for two compatible courses of action that, though relevant to the context, are distinctly independent from each other.

2. The two alternative relations are based on the use of antonymous knowledge configurations. It should be noted that antonyms should not be considered "lexical" antonyms as defined in a dictionary; rather, they are situational and contextual opposites, e.g. opposed roles or courses of actions. The following examples from the corpora:

[6.83] Should the consumer applaud its [OPEC's] efforts or fight them?
(DTel, 23/1/83, X48, 7640-1)

[6.84] 'immā jā'at tilka al-ru'ā ṣādiqatan wa nāfi<atan, wa 'immā jā'at bāṭilatan lā tanfa<u 'aḥadan.
[Either those visions are truthful and useful or they are untrue and have no benefit for anyone]
(Hr, 24/4/83, X53, 9882-3)

In both examples, the propositions represent antonyms that are arranged in a symmetrical manner for greater rhetorical effect.

A relevant pattern involves a switching of polarity of the same proposition: the first proposition is positive, the second is its negative counterpart. Rhetorically, this pattern displays an immediate opposition between alternatives.

[6.85] The Soviet regime may or may not be revisionist.
(G, 17/1/83, X10, 1470-1)

[6.86] ..taqūlu bi-al-fami al-malyāni: yajūzu, 'aw lā yajūzu!
[... they (Arab governments) say clearly: this may or may not be possible.]
(Nb, 15/7/83, X22, 4876-7)

3. The two propositions are related in semantic content to each other. This can take a number of sub-patterns. We have identified

four in both corpora (distributed differently):

a. The second proposition is an amplified version of the first one:

[6.87] ... where American nuclear weapons based on this country are concerned [Britain's] sovereignty is negotiable, or can be dismissed as almost irrelevant.
(G, 7/6/83, X39, 6314-6)

[6.88] 'iḏā qallat 'aw 'in<adamat al-maṣādiru al-'uxrā.
[if other sources are reduced or made non-existent]
(Hr, 4/1/83, X23, 4952-3)

Note that or in this pattern is supported by other connectives such as even or indeed in English, "ḥattā" in Arabic.

b. This is the reverse of the pattern in (a) above. The second proposition is a toning down of the force of the semantic content in the first proposition. Examine this example (which refers to Michael Foot's leadership):

[6.89] If Labour wins outright or obtains the largest number of seat or even if the Conservatives fail to gain an absolute majority, life begins again at 70. (O, 8/5/83, X63, 10243-6)

There is a downgrading in the way the propositions are linearised, which is supported by the connective "even if". The following example is in Arabic:

[6.90] wa qad yudammiru ḏāta-hu 'aw yabqā muḍṭariban ragma al-ṭarā'i...
[And (man) may ruin himself or stay confused despite wealth ...]
(Sh, 20/4/83, X123, 23812-3)

c. The second proposition involves specificity of the knowledge configuration expressed in the first one. The alternative relation here is weak; the two propositions convey similar content, the first is a general version while the second represents one relevant

aspect, detail or instance, or even one specific version of it. The alternation can therefore be considered a correction, or a more accurate rendering of a statement. For example, in this excerpt

[6.91] He died (or possibly shot himself) after a furious row with Suslov ...
(STi, 24/4/83, X238, 37311-37312)

the second alternative is a more specific course of action than the first one. Similarly in Arabic

[6.92] ..al-'ūbīk qad faqadat 'aw kādat tafqidu al-sayṭarata <alā sūqi al-nafṭi al-<ālamiyyi..
[OPEC has lost or has almost lost control on the international oil market ...]
(Ar, 10/3/83, X191, 33129-30)

the second alternative is a more accurate rendering of the first statement.

d. A pattern that is related to, but more intensified than, the previous one is when the second alternative statement explains or reformulates the first one. Semantically, the function is still mainly alternative, though one can argue that it also serves a clarificatory-reformulatory function (see 6.7 below).

[6.93] ... never mind what the customers want or are prepared to pay.
(O, 8/5/83, X60, 9798-9)

[6.94] lā yumkinu li-'amīrkā 'an tabniya siyāsata-hā, 'aw tuqaddima mašārī<a-hā al-silmiyyata li-miṯli hāḏā al-naw<i min "al-'umami"..
[America cannot build its peaceful policy or put forward her peaceful project for this type of nations ...]
(Nb, 15/7/83, X22, 4452-4)

In Arabic this pattern is made more forceful to the extent that the function becomes more additive than alternative (see discussion in Chapters 9 and 10). The two propositions duplicate the semantic content of each other but in different wording to create parallelistic forms. In such cases, the alternative connective

"'aw" can easily be replaced by the additive connective "wa".

[6.95] ..wa kawnu al-wāqi<i yumkinu 'an yugniya-hā 'aw yu<ammiqa min mafahīmi-hā,
[... and the fact that the current state of affairs can enrich it (Socialism) or deepen its concepts ...] (Th, 16/10/82, X66, 11993-4)

Such a pattern has been considered additive (see discussion in 6.4.8).

A final remark is made on the textual effect of negation on the alternative relation and the connectives that signal it. Examine these excerpts:

[6.96] Certainly the administration with which I was connected between 1974 and 1979 never made any such proposals or had any such intentions.
(G, 28/3/83, X17, 2698-2701)

[6.97] yajibu 'allā namalla min su'āli al ṭifli wa lā nataharraba min-hu 'aw nuḥāwila talfīqa al-jawābi.
[... we must not get bored with the child's question and must not evade it or try to make up an answer.]
(J, 5/2/83, X84, 16592-4)

Negation of alternatives means none of them is valid in the textual world. This appears to violate one of the main textual rules for the setting up of the alternative function, i.e. that one alternative should be valid. Accordingly, negation weakens alternation and in cases cancels it. In the above two examples the alternative relation expressed by or or "'aw" is weak to the extent it can be considered additive. Indeed, in English negated alternatives are often signalled via additive connectives: "nor", "neither/nor". In Arabic such additive connectives are non-existent, making negated alternation more common. Although connectives involved in negated alternation are still tagged as alternatives in the corpora, their additive meaning has to be noted.

6.7 Reformulation

6.7.1 General Comment on Textual Functioning

Reformulatory connectives indicate that the next knowledge configuration is an explication, either by way of restatement or exemplification, of the current one. This type of function is often resorted to in order to avoid ambiguity, to redefine or illustrate the informational content of a proposition. It can be regarded as a type of paraphrase whereby the possibilities of identifying the meaning of a proposition (the antecedent) are narrowed down to a single one, the one that is the centre of the text producer's concern. Accordingly, this function helps conserve substantial processing resources, and thus reduces cognitive load.

In using reformulatory connectives, the text producer adjusts text organisation via regressive sequencing (see 4.6.3.2 above). This means that the next propositions "look back" to the informational content of the antecedent. But it is a special type of regression in which the next statements enter in a conceptual loop, so to speak, with the antecedent. This may give the impression of retarding or impeding the development of content, which is true in so far as content development is temporarily delayed or upheld. But it is normally a necessary procedure, and, as with all loops, it needs to exhaust possibilities before it can be terminated.

Reformulatory connectives can be subcategorised into three: those that signal opposition and restatement of propositional content, those that illustrate it by enumerating conceptual instances, and those that restate the content in a compressed form. These will be discussed in turn.

6.7.2 Restatement

6.7.2.1 Repertoire

The following list comprises connectives that indicate restatement of text portions:

A. English

Simple: i.e., namely, plainly, rather.

Compound: in a sense, in a real sense, in one sense, in other words, in simple terms, more accurately, more pointedly, or rather, that is to say, to put (it differently).

B. Arabic

Simple: 'ay, 'a<nī, 'aw, ya<nī, bal

Compound: bi-ma<nā, bi-<ibāratin 'uxrā, bi-al-taḥdīdi, 'alā wa, ma<nā ḏālika, ya<nī hāḏā, bi-al-'aḥrā, bi-al-'aṣaḥḥi, bi-kalimatin, bi-taḥdīdin 'adaqqa, bi-ta<bīrin 'adaqqa, bi-hāḏā al-ma<nā, <alā al-'aṣaḥḥi

6.7.2.2 Textual Functioning

Connectives of this category indicate restatement of the content of a previous proposition. This can be done either by a direct paraphrase of the statement, i.e. retaining the conceptual content but in different surface text expression or by creating a more specific version of the proposition where multiplicity of meaning of a particular proposition (or a group of propositions) is reformatted to a single one. In both cases redundancy is minimised because the function, in fact, enriches, and not merely echoes, context. Even if there is some degree of redundancy, which is normally present in

all forms of paraphrase anyway, it is, in mature text production, kept within a tolerable limit. It is, after all, the text producer who has to decide whether the context justifies restatement and reformulation.

6.7.2.3 Some Textual Patterns

Two patterns can be identified in the use of reformulatory connectives, depending on the textual role of the subsequent. As with all other patterns their distribution varies in both text corpora.

1. Apposition: The connective points to some kind of appositional relationship between two (or more) propositions. This means that the subsequent proposition can stand for the antecedent. If the apposition represents a number of concepts within the antecedent proposition, or more than one proposition, the subsequent sums them up.

[6.98] Soviet communism has abolished representative government, judicial independence, freedom of association and freedom of expression: in other words it has abolished the institutions which generate legitimate opposition.
(G, 28/3/83, X20, 3109-3113)

[6.99] laqad rafaḍ-nā fī mu'tamari al-qimmati bi-al-xurṭūmi mā lā yumkinu qabūlu-hu... lākinna-nā... lam nuwaḥḥid ṣufūfa-nā wa lam nuxaṭṭiṭ li-mujābahati al-<aduwwi; 'ay 'anna-nā rafaḍ-nā dūna 'an nu<idda al-badīla.
[We rejected in the summit conference in Khartoum whatever could not be accepted ... but ... did not unify ourselves or plan for a confrontation with the enemy; that is, we made rejections without preparing an alternative.]
(Sh, 8/3/83, X117, 22578-83)

Often opposition can take the form of an explication or expansion of a concept or an informational unit within the antecedent

proposition. Note how in [6.100] the apposition helps to explain and clarify the concept of "deficiencies" stated in the antecedent.

[6.100] [The report] reveals the usual deficiencies in both sides of the British industrial relations scene - ie managements trying to foist new technology on workers without consultation, and unions, which in some cases took no steps to inform themselves about innovations likely to affect their members.
(O, 8/5/83, X61, 9921-7)

[6.101] wa yattaḍihu min ḏālika 'anna munaẓẓamata al-taḥrīri al-filasṭīniyyati baqiyat muxliṣatan li-mabda'i-hā al-'asāsiyyi, 'ay al-jam<u bayna al-bunduqiyyati wa guṣni al-zaytūni.
[It is obvious from all this that the PLO has remained faithful to its fundamental principle, that is: combining the rifle with the olive branch.]
(Sh, 8/3/83, X117, 22498-501)

2. Rephrasing: In the second pattern the consequent proposition reformulates the antecedent in order to a) achieve a correction of view (to avoid potential misinterpretation), b) offer a redefinition or give a more accurate, inclusive or specific explanation, or c) provide an inferential statement. The following excerpts exemplify reformulation for inferential purposes:

[6.102] A majority of the public would doubtless applaud such a decision [the restoration of capital punishment]. But a minority would feel bitterly outraged. In other words, the time is not ripe for the restoration of hanging.
(STel, 26/6/83, X148, 24788-91)

[6.103] ...'anna kulla 'intikāsati-hā al-sābiqati mā kānat li-taḥduṯa law kānat muwaḥḥadata al-'irādati, muwaḥḥadata al-xuṭwati, muwaḥḥadata al-hadafi. 'ay 'anna al-maṭlūba <arabiyyan, fī ẓilli fahmi al-zurūfi al-rāhinati, al-barhanatu <alā 'anna ḥālata al-tajzi'ati hiya sababu kulli al-wahni..
[... that the misfortunes of the Arab nation would never have taken place if it was united in will, step and aim. That is to say, what is needed in the present context is a proof that the current

state of dividedness is the reason behind all weakness ...]
(Th, 16/10/82, X66, 12116-21)

6.7.3 Exemplification

6.7.3.1 Repertoire

The following list of connectives introduces exemplification:

A. English

Compound: for example, for instance, for one example, such as, to take one example.

B. Arabic

Simple: maṯalan, wa,ka'an

Compound: <alā sabīli al-miṯāli, bi-mā-fī, min-ḍimni, bi-mā, ka-maṯālin, li-al-tamṯīli <alā, li-naḍrib maṯalan, min 'amṯilati (ḏālika)

6.7.3.2 Textual Functioning

Connectives that signal this relation are still appositive in nature, but it is a rather different type of apposition. The consequent proposition represents one instance, type, aspect or facet of the propositional content of the antecedent(s). Normally one representative instance is given which is sufficient to illustrate, clarify or disambiguate, thus helping to ease processing load.

6.7.3.3 Some Textual Patterns

Two textual patterns for exemplification are identified, and connectives can have any of them.

1. Illustration: In this pattern connectives signal exemplification that is intended to depict some details of the propositional content of the antecedent. Examples:

[6.104] Mr. Healey is the inventor of some memorable choice phrases. He has, for instance, taken to calling Mr. Roy Jenkins "the later Prime Minister-designate" ...
(STel, 5/6/83, X141, 23482-4)

[6.105] ..ḏālika 'anna wasā'ila targībi al-jamāhiri fī al-ṯaqāfati wasā'ilu muta<aḍḍidatun wa mutanawwi<atun. fa maṯalan yastaṭī<u al-musalsalu al-tilifizyūniyyu... 'an yaḥtawiya <alā madmūnin fikriyyin wa ṯaqāfiyyin..
[that is because the means for bringing culture to people are varied. For example, a television series ... can have an intellectual or educational context ...]
(Hr, 23/5/83, X56, 10379-84)

2. Substantiation: In argumentative texts, exemplification is often used to substantiate an argument or offer a proof. Rhetorically, this method, when used efficiently, can achieve persuasion. This pattern is particularly common in Arabic. Examples:

[6.106] Voters no longer swing the same way in all constituencies. In the case of the SDP challenge for Labour, for example, Conservatives could benefit.
(DTel, 3/6/83, X115, 19227-9)

[6.107] ..'inna-hā kulla-hā ḥurūbun murtabiṭatun ba<ḍu-hā bi-ba<ḍin, wa tan<akisu nata'iju kulli ḥarbin min-hā <alā bāqi al-ḥurūbi. fa hurūbu 'amīrka al-wusṭā maṯalan murtabiṭatun bi-ḥurūbi al-šarqi al-'awsaṭi al-qā'imati bayna al-<arabi wa 'isrā'ila,
[These wars are all related to each other, and the consequences of one reflect themselves on the rest (of the wars). For example, the wars in Central America are related to the wars in the Middle East between the Arabs and Israel ...]
(Sh, 31/5/83, X131, 25691-5)

6.7.4 Summary

6.7.4.1 Repertoire

Connectives of this category comprise a short list in both English and Arabic.

A. English

Simple: briefly

Compound: in all, in brief, in short

B. Arabic

Simple: 'ijmālan

Compound: bi-'ixtiṣārin, bi-kalimatin muxtaṣaratin, fī al-'ijmāli

6.7.4.2 Textual Functioning

Connectives of this category introduce a restatement in a condensed format of a set of earlier propositions. Usually such a restatement is made at the end of a phase in the organisational structure of the text, whatever shape that phase may correspond to, e.g. a paragraph or a bigger or smaller stretch of text. Alternatively, it may occur at the beginning of a new phase to act as a starting-point for organising the subsequent stretch of text.

In terms of information processing, a connective indicating a summary signals the initiation of procedures for re-utilising earlier knowledge blocks from mental storage. This operation involves activating earlier propositional content by bringing the contents of conceptual memory into active storage to be matched. Hence the relationship between the summary itself and earlier more detailed configuration is based on a match of underlying conceptual patterns.

It should be noted that statement of summary may depress informativity unless strong textual motivation is present. One main motivation is rhetorical: to consolidate early arguments by reintroducing them in a compressed form. It is therefore strategically sound in text not to create substantial distance between the summary statement and its more detailed version.

The following two excerpts exemplify this function of connectives in English and Arabic.

[6.108] Mr Aliyev was a renowned specialist in excessive flattery and nosing out corruption in his native Azerbaijan Republic, and now he is at the centre he will find plenty of opportunity to exercise his talents.

In short he could be the man who is going to do Andropov's dirty work ...
(DTel, 1/12/82, X94, 15930-7)

[6.109] 'inna al-<awdata li-'isdāri al-majallāti ta<bīrun <an al-ḍagṭi al-ṣarīḥi 'aw al-maktūmi li-'ifsāḥi al-majāli li-al-xalqi wa al-tafkīri wa al-ḥiwāri. wa hāḏihi al-<awdatu 'ayḍan ta<bīrun muzdawajun <an rafḍi al-ṣamti al-qā'imi wa rafḍi al-kalāmi al-sā'idi... hiya bi-'ixtiṣārin ta<bīrun <an al-ḥājati 'ilā ṭaqāfatin jadīdatin.
[The return to publish these journals is an indication of a pressure, both overt and covert, for providing opportunities for creation, reflection and interaction. This return is also a double expression of rejection: rejection of current silence and current talk ... It is, in short, an expression of the need for a new type of culture.]
(Sh, 22/2/83, X111, 21494-503)

6.8 Orientative

6.8.1 General Comments on Textual Functioning

Connectives of this category signal a modified or revised point of view. The text producer, after introducing a view or comment, ostensibly detaches himself from his sequence of arguments and takes

an authoritative stance (either objective or subjective) from which he makes judgements, modifications of attribution, or enlarges his argument so that ambiguities or misapprehensions are corrected, or moves to a new stage in the exposition. In terms of processing, the connectives, generally speaking, signal a turning point where a path is initiated to enable the processor to track back and reprocess the data. Or, the connective can signal a stage where the processor, after moving along a track, has to calculate competing alternatives and selects one. In this case the connective assists the processor in making a sharper grasp of relative possibilities by modifying the knowledge of context and world situation.

Connectives of this category may be categorised according to their textual role into two types. The first is view-adjustment, where viewpoints are modified in some way; the second is confirmation and reinforcement where the content of a proposition is specifically emphasised. Those will be discussed in detail next.

It should be brought into attention that connectives of this category are different from those of other categories in two ways:

1. The repertoire of this category, compared to the previous ones, comprises an open class of expressions. This is indicated by the big number of types within this category and the variation of functionality that they signal. This openness is rhetorically essential since it enables the text producer to select those that can more accurately, appropriately or effectively adjust the perspective through which propositions are to be formulated.

2. The repertoire comprises expressions that, because of their

grammatical class, may function as textual connectives in some contexts and have a different function in others. These expressions are predominantly adverbials (in English), or prepositional phrases, "tamyīz" or absolute accusative "maf<ūl muṭlaq" (in Arabic). An expression such as certainly or frankly, for instance, may function as a disjunct acting as a textual connective; but they may occur as adjuncts and have a structural role limited to the clause. (For a detailed account of the semantic functions of the adverbials in English see Quirk et al. 1972, 1985).

6.8.2 Adjustment

6.8.2.1 Repertoire

The following list comprises connectives of this category.

A. English

Simple: allegedly, apparently, as, basically, characteristically, crucially, crudely, curiously, emotionally, especially, essentially, financially, frankly, ideally, inevitably, literally, logically, mainly, nationally, normally, now, overall, particularly, partly, personally, politically, possibly, practically, precisely, presumably, primarily, privately, significantly, simply, statistically, strategically, tactfully, theoretically, traditionally, well.

Compound: against (this, that), as far as, at first sign, at least, at a deeper [superficial] level, at one [some] point, by definition, by all accounts, by and large, economically speaking, for me [them],

for once, for one thing, for its [my, my own] part, for that matter, from (his) point of view, in effect, in general, in general terms, in part, in particular, in parts, in practice, in private, in real life, in retrospect, in such (a crisis), in theory, in a sense, in a way, in an historical perspective, in both cases, in both respects, in his view, in his own eyes, in (immediate electoral, practical) terms, in many cases, in many ways, in my opinion, in some cases, in some respects, in this [that] case, in this [that] context, in this [that], respect, in the eyes of, in the real world, in the sense (that), in the short term, in this regard, in this role, in this sense, in this situation, in those respects, in (Tory) eyes, in which case, it seems to (me), materially speaking, more ambitiously, more crucially, more important, more seriously, more consensually, not unreasonably, on the face of it, on the whole, so far as, so to speak, through her eyes, to all appearances, to all intents and purposes, to be fair, to (everyone's, his, my, no-one's) surprise, to Western eyes.

B. Arabic

Simple: 'ammā, kamā, xaṣṣatan, xuṣūṣan, siyyamā, 'izā'a, ʿaxṣiyyan, <umūman, <inda, hunā

Compound: <alā al-'aqalli, fī hāḏā al-ṣadadi, fī hāḏā al-

majāli, fī ra'yi, min al-nāḥiyati, min ḥayṯu, fī hāḏihi al-ḥālati, min hāḏā al-munṭalaqi, fī hāḏihi al-ḥāli, fī al-maydāni, fī 'i<tiqādi, min wijhati naẓari, bi-al-ḏāti, bi-xaṣṣatin, fī naẓari, fī hāḏā al-'iṭāri, bi-al-'asāsi, bi-al-'axassi, <alā al-'axassi, fī ḍaw'i, fī 'i<tiqād, fī 'iṭari, fī taqdīr, fī hāḏā al-maqāmi, fī hāḏā al-siyāqi, fī kiltā al-ḥālatayni, min al-ma<rūfi, min al-mulāḥaẓi, ḥasaba ra'yi-hi, <ilman-bi, bi-al-darajati, bi-šaklin 'awḍaḥa, bi-šaklin xāṣṣin, bi-gayri, bi-naẓari, bi-wajhin <āmmin, bi-ṣarfi al-naẓari, bi-hāḏā al-'ittijāhi, bi-hāḏā al-majāli, bi-hāḏā al-ṭarīqi, bi-hāḏihi al-munāsabati, bi-haḏihi al-tarīqati, bi-kulli ṣidqin, bi-nafsi al-mi<yāri, <alā al-ṣa<īdi (al-'iqtiṣādiyyi), <alā al-<umūmi, <alā al-mustawā (al-'iqtiṣādiyyi), <alā mustawā, fī al-'aqalli, fī al-ḥalatayni, fī ḥalati, fī 'i<tibāri, fī 'i<tiqādi-nā, fī 'ijtihād-ī, fī manḥā-hu hāḏā, fī taṣawwur-ī, fī taqdīri, fī yaqīn, fī hāḏā al-mafhūmi, fī hāḏā al-maydāni, fī hāḏā al-miḍmāri, fī hāḏā al-ṭarīqi, fī hāḏā al-sabīli, fī hāḏihi al-nuqṭati, fī miṯli hāḏihi al-ḥālati, min al-badāhati, min al-manṭiqiyyi, min al-muṯīri, min al-saḏājati, min ḥusni al-ḥaẓẓi, min bābi al-wāqi<iyyati, min <ajabin, min munṭalaqi, min hāḏā al-naw<i

6.8.2.2 Textual Functioning

This category comprises a wide variety of textual connectives. The general function that these ostensibly diverse types have in common is to enable the text producer to imply his own attitudinal presence by creating points where sequences of propositions, expressing views, states or courses of action, can be adjusted: modified, heightened, specified or generalised. This, in effect, controls the flow of the arguments and provides them with the necessary transition.

"Adjustment" is a cover term that is used here to designate transitions of several types.

1. Assigning authority: The text producer specifies whose authority it is assumed in the expression of a certain (set of) propositions. In the English corpus, connectives expressing this function are: personally, from my [his, etc.] point of view, for them, to me [them, etc.], for my [its, etc.] part, in my-opinion, in his view, in [to, through] his [her, etc.] eyes and as far as ... concerned. In the Arabic corpus, the following expressions serve as connectives signalling this function:"fī 'i<tiqādi [-nā, -ī, etc.]" [in our (my) opinion], "fī taṣawwuri [-nā, -ī, etc.]" [in our (my) view], "fī taqdīr [-i, etc.]" [in my judgement], "fī [ḥasaba] ra'yi [-hi, etc.]" [in his opinion], "fī naẓar [-ī, etc.]" [in my view], "fī yaqīn [-ī, etc.]" [in my conviction], "šaxṣiyyan" [personally], "<inda [-nā, etc.]" [To us], "min wijhati naẓari [-hi, etc.]" [from his point of view].

[6.110] In spite of certain professional malaise resulting from my sojourn in the USSR, I had every reason to be thankful for it. From my point of view it had been infinitely worth while, enabling me to

understand as nothing else would what the Soviet regime was about, how it functioned and what was its impact on neighbouring countries and the world in general.
(G, 11/4/83, X25, 4022-9)

[6.111] ..'anna hādā al-ṣira<a al dā'ira ḥawla taḥsīni tilka al-mubādarati 'aw al-ṣiyagi yumaṭṭilu fī ra'yi-nā ṭabī<ata al-ṣirā<i al-<arabiyyi al-'isrā'iliyyi al-dā'iri fī hāḍihi al-marḥalati allatī nastaṭī<u 'an nuṭliqa <alay-hā "marḥalata manāhiji al-salāmi".
[This current conflict over improving these initiatives and formulae reflects, in our view, the nature of the Arab-Israeli conflict at this stage, which we can designate "Peace Programmes Stage".] (Hr, 24/1/83, X27, 5706-11)

Another, structurally different but functionally similar connective that is far more frequent within this category, both in the English and Arabic corpora, is as and its counterpart "kamā" Assignment of authority via these connectives can occur either before or after the statements are made; but most frequently, particularly in Arabic, it is parenthetical, i.e. it occurs within the statements themselves. In this latter case, the connective can create salience by interrupting a statement to assign its source of reference.

Observation of the functionality of these connectives suggests three main types with possible variation in each.

a. Type (i): In this type there is a clear indication of the authority to which statements are attributed. Some examples from the English corpus:

as Lord Young says,
as one of them put it,
as the BIS report explains,
as the researchers have been telling us,
as Edwards himself has pointed out,
as the Government's critics see it,
as Joseph Berliner has argued, etc.

The following examples are from the Arabic corpus:

kamā <abbara <an ḏālika al-ra'īsu al-qā'idu
[as stated clearly by the president]

kamā 'ašāra 'ilā ḏālika taqrīru al-'umami al-muttaḥidati
[as a UN report pointed out],

kamā yaqūlu al-mutaḥadditu al-<askariyyu
[as the military spokesman says],

kamā tuṣirru 'isrā'īlu
[as Israel insists],

kamā na<taqidu
[as we believe].

b. Type (ii): In this type the attribution of source of reference is not clear and is, hence, indirectly stated. The connective serves to assess the status of the authority rather than identify it, and can reflect the text producer's attitude towards the statements being made. This, in effect, dictates some variation in the way attitudes are specified. The following are examples from the English corpus:

as it was known,
as was predicted,
as it is,
asit happened,
as it well may,
as widely sensed,
as things now stand,
as it seemed, etc.

The following examples are drawn from the Arabic corpus:

kamā yabdū [as its seems],
kamā yattaḍiḥu [as it gets obvious],
kamā huwa ma<rūfun [as it is well-known],
kamā yalūḥu [asitlooks],
kamā zu<ima [as it has been claimed],
kamā tafriḍu-hu al-naẓratu al-wāqi<iyyatu
[as dictated by practical attitudes].

2. Assessment: This refers to the nature of the authority that is assumed and reflects its content. Connectives of this function enable the text producers to make an assessment or judgement of the present status of his views, comments, etc., and hence signal large and diverse types of meaning. Examples from the English corpus are: allegedly, apparently, basically, characteristically, crucially, crudely, curiously, emotionally, frankly, logically, partly, privately, significantly, tactfully, at least, in private, to be fair, to my [his, etc.] surprise, etc.

In the Arabic corpus, connectives of this category represent a comparatively smaller number of expressions:

"min al-badāhati", "min al-badīhiyyi" [intuitively, apparently], "bi-al-'asāsi" [basically], "min ḥusni al-ḥaẓẓi" [fortunately], "min al-ma<rūfi" [evidently or generally accepted], "min al-saḏājati" [naively enough], "min <ajabin" [strangely enough].

The following excerpts exemplify the function:

[6.112] ... the abolition of the 25 per cent tax band caught many low income groups in the net. Logically in order further to improve the generous incentives for new businesses and encourage employment, the upper rates of tax should be increased again. (G, 14/4/83, X26, 4323-8)

[6.113] ... let me float an alternative, perhaps interim, idea that was suggested to me by a young first-time voter. It would at least deal with the problem of tactical voting.
(T, 26/6/83), X227, 35721-4)

[6.114] fa al-ṯawrātu bi-al-'asāsi <amaliyyatu taṣaddin šujā<atun li-al-bunā al-mutaxallifati bi-mā fi-hā al-taqālidu al-bāliyatu..
[For the revolution, basically, is a brave operation of counteraction of backward life patterns including obsolete traditions ...]
(Th, 5.4.83, X73, 13704-6)

3. Generalisation: The text producer may, on the basis of the content of the previous text sequence, wish to generalise a state, view or attitude. This is an important step in the design of a text and represents the text producers' effort to extract from various individual comments, views or instances a more elaborate one that can more effectively represent his discussion or point of view. In the English corpus connectives that express this function are: overall, in general, on the whole. In the Arabic corpus, such expressions include: "<umūman" [generally], "bi-wajhin <āmmin" [in general]. For example:

[6.115] Zeldin's purpose is to discern ... the truth behind the image or stereotype of the French. He is very knowledgeable and perceptive, and anyone who loves, or admires, or for that matter hates, the French will find this book a delight.
On the whole it serves to undermine prejudices.
(DTel, 29/1/83, X96, 16264-71)

[6.116] 'inna al-tasā'ula yanhaṣiru fī ma<rifati sababi tadaxxuli al-mušarri<i al-jina'iyyi fī qanūni al-<uqūbāti al-miṣriyyi bi-tajrīmi ba<ḍi 'af<ālin taqūmu bi-hā al-ṣuḥufu wa gayru-hā min wasā'ili al-'i<lāmi bi-ṣadadi al-našri <an al-muḥakāmati wa al-'ahkāmi. wa bi-wajhin-<āmmin fa 'inna ragbata al-mušarri<i laysat 'illā kabḥa jimāḥi al-ṣihāfati 'iḍā xarajat <an wājibi-hā al-muqaddasi..
[The question confines itself to understanding the reason behind the Egyptian penal code by incriminating certain courses of action that the press and other types of media have taken when they propagate court-proceedings and sentences. In general, this step of the legislator is intended only to restrain the press if it deviates from its sacred duty ...]
(Hr, 23/5/83, X55, 10146-53)

4. Particularisation: Connectives expressing this function enable the text producer to narrow down the content of a previous sequence of propositions. This helps shift focus from the more general to

the more specific and creates space for more textual development. In the English corpus expressions that function as connectives of this category are: especially, particularly and in particular. Their counterparts in the Arabic corpus are "xāṣṣatan", "<alā al-'axaṣṣi", "xuṣūṣan", "siyyamā", "bi-xāṣṣatin", and "bi-al-ḏāti", all having the meaning of "particularly" or "in particular". Examples:

[6.117] The Camp David accords of 1978, trumpeted as the blue-print of the global settlement, were allowed to decline into a separate Egypt-Israel peace. Israel was freed to use its muscle on its eastern frontier. In particular, America has failed to check Israeli settlements on the West Bank.
(O, 26/6/83, X92, 15521-6)

[6.118] wa al-qaḍiyyatu laysat jadīdatan <alā al-naqdi. fa qad 'utīrat <ašarātin min al-marrāti wa xāṣṣatan ba<da 'an naẓama šawqi masraḥiyyata-hu "maṣra<u kilyubātra"...
[This case is not new to literary criticism. It was raised tens of times, particularly after Shawqi wrote his play "Cleopatra's Death"]
(Hr, 4/3/83, X23, 4961-4)

5. Variation in viewpoint: In the English corpus this function is associated with the connectives against that, as for and incidentally. In the Arabic corpus, it is a more frequent function, signalled by the connective "'ammā". Other Arabic connectives of this category are: "'izā'a" [against (this, that)], "hunā" [at this, here], "bi-hādihi al-munāsabati" [incidentally]. These connectives indicate that a variant viewpoint is to be introduced, that, related to a previous one, represents an opposition, an interruption or a point of departure. Textually, the connectives help establish transitional points that enable the text producer to introduce various facets of his arguments. In this excerpt, for example

[6.119] Against that background, it seems reasonable to run a beef enterprise and plan as carefully as any other ...
(DTel, 18/4/83, X104, 17381-3)

"that background" refers to the previous six paragraphs where questions related to "beef production" are discussed. Note that in the case of incidentally, the flow of the statements is interrupted to introduce a diversion: either to comment on, define or enrich the content of the previous stretch of text.

In Arabic the connective "'ammā" normally requires the connective "fa" in an additive/commentative meaning to support its textual function. For example:

[6.120] wa ka-al-ḥāwi alladī lā yahummu-hu min al-li<bati 'illā al-juz'u alladī yakšifu <an-hu li-jumhūri al-mutafarrijīna, 'axraja jūrj šūltz min jirābi-hi al-diblūmāsiyyi mašrū<a 'ittifāqiyyati salāmin bayna lubnāna wa 'isrā'īla, laysa fī-hi 'illā al-haykalu al-xārijiyyu; 'ammā al-maḍmūnu ma<a al-malāḥiqi, ma<a al-tafsīrāti, ma<a al-murāsalāti al-sirriyyati al-mutabādalati, fa qad ta<ammada jūrj šūltz 'an yubqiya-hā kulla-hā maxfiyyatan fī dāxili "jirābi hi" al-mashūri..
[And as a magician whose main interest in the show game is the part that he exposes to the audience, George Shultz took from his diplomatic hat the draft of a peace settlement between the Lebanon and Israel, that has only an external structure. As for the content including the appendices, the interpretations, the secret correspondences: George Shultz had decided to keep them in his magic hat.]
(Nb, 20/5/83, X9, 1046-55)

Sometimes, the opposition or divergence as expressed by these connectives is so sharp that the general meaning conveyed is more adversative than orientative. In such instances the connectives have been classified as adversative/contrastive.

6. Specification of Perspective: Connectives of this category specify the angle, perspective or point of view from which a statement is to be considered. Like connectives of category 5 above (i.e. variation of viewpoint), but without the characterising rebutting or divergence in the statement of arguments, connectives of specification establish transitional points that may lead to either one of two types of orientation paths:

a) Type (i): a new perspective, different from the one in which the content of the current or previous sequence is viewed. The following expressions are examples of English connectives of this type:

economically speaking,
financially,
politically,
nationally,
statistically,
strategically,
at one point,
at a superficial level,
in real life,
in retrospect,
in theory,
in a historical perspective, etc.

Arabic connectives of the same type are:

fī ḥālati [in the case of],
fi al-maydāni (al-'ijtimā<iyyi, al-'iqtiṣādiyyi, al-ṯaqāfiyyi) [on the (economic, social, cultural) level],
<alā al-mustawā al-<arabiyyi [on the pan-Arab level],
<alā al-ṣa<idi al-'iqtiṣādiyyi [on the economic level].

The following two excerpts exemplify this function:

[6.121] Much the same is true of Thatcherism on the domestic front. There, too, the driving force is moralistic rather than materialistic. At a superficial level the intention is, of course, to restore economic growth and sound money. But at a deeper level, as is widely sensed, it is to restore moral growth and sound values.
(STel, 1/5/83, X129, 21479-85)

[6.122] wa fī al-maydāni al-'iqtiṣādiyyi kāna banku al-tanmiyati al-'islāmiyyi ṭalī<ata al-munjazāti al-kubrā allatī ḥaqqaqat-hā al-munaẓẓamatu. wa huwa yaqūmu mundu 'inšā'i-hi sanata 1975m bi-dawrin našīṭin fī musā<adati al-duwali al-'islāmiyyati fī tanfiḍi mašrū<ati-hā al-tanmawiyyati..
[On the economic level, the Islamic Development Bank was one of the great leading achievements that the Organisation [Islamic Conference Organisation] has produced. Since it was established in 1975, the bank has played an active role in helping Islamic countries in carrying out their programmes of development ...]
(Sh, 22/1/83, X107, 20755-60)

b) Type (ii): Orientation is restricted to a perspective that is identified within the current or previous sequence of text. In this case, the connective is a multi-word expression containing an anaphoric referential item that points to the text sequence that represents the perspective or points of view. Examples of such connectives in English are:

in this respect,
in this regard,
in this case,
in those respects,
in this situation,
in this context,
in this role, etc.

Examples of Arabic connectives of the same type:

fī hāḍā al-ḥāli [in this case],
fī hāḍihi al-nuqṭati [in this point],
fī hāḍā al-siyāqi [in this context],
fī hāḍā al-maydāni [in this level],
fī hāḍā al-ṣadadi [in this connection],
fī hāḍā al-'iṭāri [within this framework or context], etc.

The following excerpts exemplify the use of this type of connectives.

[6.123] There are very few fields indeed, in the arts, sciences, technology, education, transport, institutions or the sheer business of living well and gracefully, in which the French are not accomplished and innovatory performers, very near the top of the world league. In that respect, no other race can match them ...
(DTel, 29/1/83, X96, 16242-8)

[6.124] fa lam taṭraḥ al-qimmatu al-sābi<atu mubādaratan jadīdatan tatalāfā al-ṣu<ūbāti al-<amaliyyata allatī tuwājihu mašrū<a fās wa 'in kānat qad ḥāwalat al-musāhamata fī taḥrīkin jadīdin yu<ṭī duf<atan li-hāḏā al-mašrū<i. wa fī-hāḏā-al-'iṭāri jā'a taškīlu lajnatin min 7 ru'asā'a min 'a<ḍā'i al-mu'tamari li-mutāba<ati tanfīḏi al-qarārāti al-xāṣṣati bi-al-qaḍiyyati al-filastīniyyati..
[The seventh Summit [of the Non-aligned movement] did not introduce a new initiative that could avoid the practical difficulties facing the Fez initiative, though it (the Summit) tried to participate in setting this initiative in motion and giving it further backing. Within this framework, a committee was formed comprising 7 heads of states who are members of the Movement to follow up the implementation of the resolutions that concern the Palestinian cause ...]
(Hr, 30/3/83, X40, 7389-96)

6.8.3 Confirmation

6.8.3.1 Repertoire

The following expressions may function as connectives of confirmation:

A. English

Simple: actually, certainly, clearly, doubtless, evidently, indeed, invariably, naturally, obviously, rightly, surely, true, undoubtedly, unmistakably, yes.

Compound: as a matter of fact, in fact, in reality, in truth, needless to say, no doubt, no less importantly, not surprisingly, of course, to be sure.

B. Arabic

Simple: fi<lan 'innama, na<am, haqqan, tab<an, hasanan, 'ajal, yaqinan, badahatan, kalla, tayyib.

Compound: la šakka, bi-al-ṭab<i, bi-al-fi<li, bi-al-ta'kīdi, fi al-wāqi<i, fi al-ḥaqīqati, bi-lā šakkin, min al-wāḍiḥi, bi-dūni šakkin, bi-ṭabī<ati al-ḥāli, min

al-mu'akkadi, bi-al-qaṭ<i, fi wāqi<i, min al-ṭabī<iyyi, sawā'an bi-sawā'in, bi-kulli al-maqāyīsi, bi-lā 'adnā šakkin, lā rayba, li-al-ḥaqīqati, mā min šakkin, min al-badīhiyyi

6.8.3.2 Textual Functioning

Occasionally the text producer needs to confirm the content of a text stretch and to create forceful assertions. There are various structural means of achieving this effect, one of them is the use of the expressions that we here designate "connectives of confirmation".

Generally speaking, connectives of this type introduce a reinforcing effect to the content of the proposition to which they are attached, or, occasionally, to the content of the current or previous propositions. Their function is to confirm the truth value of a proposition, to underlie the certainty of its content or, at least, to increase the possibility of an "untrue" statement being taken as the truth.

This function is occasionally conflated with one (or more) of three other types of function, thus giving three subcategories: confirmation proper, conclusive, and corrective. This classification is based on the relation that the connective establishes as it connects the subsequent proposition with the current or previous one. In a number of cases, we have admittedly experienced some difficulty in drawing a demarcating line between these subcategories: the functions seem to merge together. However, the classification is necessary to help understand the functional nature of these connectives and their rhetorical impact in the text.

1. Confirmation proper: Connectives having this function are mainly concerned with confirming the propositional content of the statement (or sequence of statements) they apply to. Strictly speaking, expressions such as yes or, in Arabic, "'ajal" functioning as connectives of this category, normally provide confirmation to the previous text stretch, while such expressions as certainly, undoubtedly, and of course, or in Arabic, "fi<lan", "haqqan" and "bi-dūni šakkin", tend to apply to the current or subsequent statements. However, in both cases, confirmation works in both directions in the text; the subsequent statement provides support to the current or previous one, thus achieving an integrating cohesive effect. The following excerpts exemplify this function.

[6.125] Mr. Buckle also says the company has been attacking free speech. Certainly senior stewards have been told to stop criticising BL in public although many have ignored this warning.
(G, 20/4/83, X27, 4501-4)

[6.126] ..lākinna-nī 'arad-tu al-tarkīza <alā nuqṭatin 'awaddu munāqašata-hā, wa hiya fi<lan mā da<ā-nī li-al-ta<līqi <alā hāḏā al-mawḍū<i.
[... but I wished to concentrate on an important issue that I would like to discuss next and it is indeed what made me comment on the whole subject.]
(Nb, 5/4/83, X6, 712-5)

While certainly and "fi<lan" confirm the content of the statements they introduce, they, simultaneously, provide confirmation to the truth value of the previous statements.

2. Conclusive confirmation: Connectives of this category, while providing confirmation to the content of the subsequent proposition (the one to which they are attached), signal a conclusion to the current or previous one(s). The conclusion is usually based on

evidence presented explicitly or implicitly within the current or previous stretch of text, and the existence of the connective enables the text producer to assert his prediction. This is exemplified in the following excerpts:

[6.127] Mrs. Thatcher sometimes gives the impression that gaps are things to leap into and widen as much as you can with your bare hands and then say briskly: "Right. That is my side of the gap. Over there is the wrong side. Make up your mind which side you're on." Clearly, as conventional political style, it breaks the rules.
(DTel, 1/6/83, X113, 18878-85)

Here, "clearly" while confirming the statement it is attached to, introduces a comment formulated on the basis of the evidence given in the earlier stretch of text. Similarly in Arabic.

[6.128] kamā 'anna hāḏā al-sayla al-<ārima min al-ma<lūmāti badalan min 'an yakūna maṣdara xayrin wa hudā li-al-fardi 'aṣbaḥa maṣdara taḍlilin. wa bi-al-ṭab<i fa 'inna al-ṣawta al-'aqwa wa al-'awḍaḥa fi hāḏihi al-fawḍā al-ḍāribati yakūnu li-man yamtaliku al-'imkānāti al-kabīrata. wa man yamtaliku hāḏihi al-'imkānāti gayru al-duwali al-mutaṭawwirati?
[Besides, this powerful flow of information [through the media], instead of being a source of goodwill and guidance to the individual, has become a source of misguidance. And definitely the strongest and clearest voice belongs to the nation that has ample resources. And who has these except the developed countries?]
(J, 10/4/83, X93, 17854-61)

the writer, by using "bi-al-tab<i" [definitely, naturally], offers a view concluded on the grounds of his arguments in the earlier stretch of text.

3. Corrective Confirmation: Similar to the previous type, connectives of this category mark a conclusion based on an evidence presented within the current or earlier stretch of text. The

conclusion has the additional function of correcting another assertion that has been introduced. For instance in the following excerpt:

[6.129] Strikes are irrelevant to a company losing money. Indeed they can be welcome, for what is the harm in losing production if you cannot sell the things that you produce?
(DTel, 1/6/83, X114, 18991-4)

the connective indeed, while making a confirmation, offers a correction of view.

In Arabic, this function is rare. A text producer who would like to use corrective confirmation will use a confirmative connective preceded by an adversative connective, e.g.

lākinna-hā bi-al-ṭab<i [but definitely],
lākinna-hu bi-al-ta'kīdi [but certainly],
lākinna-hu .. bi-lā šakkin [but.. without doubt],
ragma .. fī al-wāqi<i [though ... in fact], etc.

Alternatively, an adversative/antithetic connective is used, e.g. "bal". These two alternatives also apply to English when this function is intensified to the extent where it becomes adversative. In this case a confirmative connective is either preceded by an adversative, e.g. but actually, but certainly, yet ... clearly, but obviously, etc., or alternatively the connective itself may function more as an adversative (signalling a contrast in propositional content) than a confirmative, in which case it is categorised under adversative connectives (as in the adversative use of of course).

6.9 Temporal

6.9.1 General Comments on Textual Functioning

The basic function of temporal connectives is the provision of

time references that help organise the various knowledge configurations in the textual world. This is particularly true in narrative texts or in expository texts that have a narrative component, whether a big or small one. In such texts, fine distinctions in time relationships are required to render the textual world tidy and manageable.

It is true that finite verbs can create temporal organisation that is necessary in any text; this is normally produced normally via signalling tense and aspect and by the use of prepositional phrases acting as time relators. But temporal connectives organise propositions further by ordering their informational content in terms of time location. This can serve cohesion by relating propositions to each other temporally, by organising sequencing through reinforcing progressive or regressive linearisation, and by sustaining salience through keeping temporal ordering in focus.

Temporal connectives that are identified in the two corpora signal six subcategories of temporality, reflecting types of ordering of events or states across time. These are: 1) temporal sequence, 2) simultaneity, 3) span, 4) temporal positioning or punctiliarity, 5) temporal frequency, and 6) circumstance.

6.9.2 Sequence

6.9.2.1 Repertoire

The following connectives signal the relation of temporal sequence:

A. English

Simple: after, afterwards, and, before, eventually, finally, first, initially, later, next, originally,

subsequently, then, ultimately.

Compound: at first, at last, at long last, at the end, in retrospect, in turn, in the end, in the first place

B. Arabic

Simple: wa, ba<da, ṯumma, fa, qabla, 'axīran, qablu, 'awwalan, 'aṣlan, 'awwala, bad'an

Compound: fī al-nihāyati, min qablu, wa min ṯamma, bi-al-tāli, fī al-bidāyati, bādi'a ḏī bad'in, fī nihāyati (al-maṭāfi), min al-bidāyati, munḏu al-bidāyati, min ba<di (ḏālika).

6.9.2.2 Textual Functioning

Temporal connectives of sequence signal succession of events, states or courses of action. This function establishes an order through which informational units follow one another on a temporal basis. It is a powerful tool for controlling text sequencing: for determining the manner in which propositions are juxtaposed, and for guiding textual progression where one proposition (as an informational unit) looks ahead to the next one in the order of succession, and textual regression where a later proposition looks backward to the previous one.

6.9.2.3 Some Textual Patterns

An examination of this function as signalled by connectives in the two corpora suggests three subcategories of temporal sequence, depending on the nature of the succession of the propositions:

a) The succession "proper" is signalled mainly by the connectives and, then, next and later in English, "wa", "fa" and "ṯumma" in

Arabic. These indicate a clear ordering of events, with the antecedent taking place on the time continuum before the subsequent. The succession may be immediate, i.e. one event takes place immediately after another, or there may be an indefinite time lag. The connectives and and "wa" can signal either. In Arabic, the connective "fa" normally signals immediate succession only. It, therefore, displays more specificity of functioning than and or "wa". The following are examples of immediate succession (where there is a short time lag).

[6.130] ... she (Princess Diana) suddenly stopped, leapt out screaming and ran away into the crowd.
(STi, 23/1/83, X233, 36716-7)

[6.131] ..jā'at al-marḥalatu al-'axīratu li-al-'insiḥābi fa ẓaharat qiṣṣatu "ṭābā".
[... the last stage of the evacuation arrived and the story of "Taba" came into existence.]
(Hr, 26/1/83, X35, 6611-2)

The connectives then, later, next and "ṯumma" usually indicate that there is a time lag separating the two events.

[6.132] Outside the Basque it [the French beret] was almost unknown until 1920. Then it became a national uniform ...
(DTel, 29/1/83, X96, 16308-10)

[6.133] hiya majallatu "Šiʿr" allatī kāna yuṣdiru-hā Yūsuf Alxāl fī Bayrūta, ṯumma tawaqqafat ʿāma 1970..
[... it is the journal "Shi'r", which was edited by Yusuf Alkhal in Beirut, and later ceased in 1970 ...]
(Sh, 22/2/83, X111, 21405-7)

The time lag can be indefinite or it can be specified by an adverbial phrase that follows the connective immediately, e.g. "Then last week" or "ṯumma baʿda sanawātin [then after a few years]". It should be noted that most other temporal connectives can signal this function.

b. Succession can be established by indicating the initial or final stages. The initial stage is indicated in English by the connectives: first, initially, originally, at first, in the first place. In Arabic it is signalled by the connectives: "'awwalan", "'awwala mā", "bad'an", "bādi'a ḏī bad'in", "fī al-bidāyati", "min [mundu] al-bidāyati". For example:

[6.134] Initially success is more likely to be achieved using egg cells of primates ...
(G, 2/12/82, X6, 892-3)

[6.135] ..taṣawwara al-ba<ḍu 'anna al-ḥarba bayna 'amīrkā wa ḥalīfāti-hā fī 'ūrūbba waš̄ikatu al-wuqū<i fī al-majāli al-'iqtiṣādiyyi 'awwalan.
[Some believed that an economical war will first break out between America and her allies in Europe ...]
(Sh, 1/3/83, X114, 21914-6)

The final stage of succession is signalled in English by the connectives: finally, eventually, ultimately, at least, at long last, at the end, in the end. In Arabic, it is indicated by the connectives: "'axīran", "fī al-nihāyati", "bi-al-tāli".

c. A chain of sequential events is achieved in Arabic by repeating the connective each time a new event is introduced. Connectives that are usually used in such chains are "wa", "fa" and "ṯumma". For example:

[6.136] 'inna Miṣra allatī <āradat fikrata al-dawlati al-filasṭīniyyati wa lam tu'ayyid jihāran nahāran 'aw bi-al-kāmili qarārāti mu'tamari "Fās", ṯumma saḥabat mašrū<a qarāri-hā al-muštaraki ma<a faransā min al-'umami al-muttaḥidati wa min majlisi al-'amni, ṯumma <āraḍat qarārāti mu'tamari "al-bunduqiyyati" li-duwali 'ūrūbbā al-<aš̄ri, ṯumma la<ibat wa ta'arjaḥat wa 'ayyadat ṯumma rafaḍat ṯumma 'ayyadat muxtalafa al-mašārī<i..
[Egypt, which objected to the notion of a

Palestinian state and did not approve the declarations of Fez conference, then withdrew its bi-lateral statement with France from the United Nations and the Security Council, then rejected the declarations of Venice conference of the ten European states, then manoeuvred, changed sides and approved, then rejected then approved various projects ...]
(Nb, 20/5/83, X9, 1251-1260)

An examination of the English corpus has suggested that such chaining is restricted. When a chain is introduced, it is usually shorter and the connective is often reserved to introduce the last proposition. For example:

[6.137] Rachel brightened, pored over the courses on offer and eventually plumped for French Provincial cookery.
(O, 8/5/83, X58, 9313-5)

6.9.3 Simultaneity

6.9.3.1 Repertoire

Connectives that signal the temporal relation of simultaneity are:

A. English

Simple: and, as, meanwhile, simultaneously, when, while

Compound: all this time, at the same time, in the interim, in the mean time.

B. Arabic

Simple: 'iḏ, baynamā, fīmā, wa

Compound: fī al-waqti nafsi-hi, fī nafsi al-waqti, fī ḏāti al-waqti, fī waqtin wāḥidin, fī <ayni al-waqti, fī al-waqti <ayni-hi, fī al-waqti ḏāti-hi, fī tilka al-'aṯnā'i.

6.9.3.2 Textual Functioning

The main function of connectives signalling temporal simultaneity is to indicate that the event, course of action or state expressed in the subsequent proposition runs parallel to that in the antecedent. In terms of information, the conceptual configurations of both propositions are made active at the same time; the connectives help the processor keep the antecedent proposition active while it accesses the subsequent. The integration of the two propositions results as a consequence of sharing the same time location or time range in the textual world.

6.9.3.3 Some Textual Patterns

In the corpora, connectives of temporal simultaneity seem to have two distinct textual patterns:

1. Simultaneity proper: In this pattern, the main focus is on the temporal coincidence of two propositions. Normally the second proposition shares either one point, a limited location or the entire range of time adopted in the first proposition.

[6.138] Among all the challenges the Labour party faces as it enters into this long dark night of prolonged opposition, the greatest is the recovery of credibility.
(Ti, 20/6/83, X226, 35616-8)

[6.139] wa baynamā hiya taḥtariqu, šā'at muṣādafatun 'an yamurra ḏālika al-rajulu min hunāka fa yašhada al-mar'ata al-qiddīsata
[And as she was burning, that man passed that way by chance ...]
(Hr, 24/4/83, X53, 9663-5)

2. "Abutting" simultaneity: In this pattern, temporal simultaneity is overshadowed by the way the two propositions

confront each other, posing two different perspectives, or expressing two different conceptual contents. The confrontation is, however, not taken to the point where the relation is adversative, though a connective such as while in English can still do that. Rather, the propositions reflect courses of actions that are unrelated except in their temporal coincidence. This abutting meaning of the pattern explains why the connective "fī al-waqti nafsi-hi" and its English counterpart at the same time are occasionally preceded respectively by the adversative connectives "wa lākin" and "but". In general, this pattern is more frequent compared to the previous one, particularly in Arabic. Examples from the corpora:

[6.140] The Russians seem, quite sensibly, to have decided that Pakistan is the weak link in the anti-Afghan coalition aid that it may respond better to smiles than frowns. At the same time, they are trying to put the blame for Afghanistan's difficulties on what they call the world-wide "anti-Soviet" manoeuvres of Washington".
(O, 26/6/83, X89, 15144-50)

[6.141] ..kādat maṭābi<u Miṣra wa Lubnāna 'an taxtaniqa bi-al-maxṭūṭāti al-jadīdati allatī tantaẓiru dawra-hā li-'axḍi makāni-hā fī al-maktabāti al-<arabiyyati. wa fī-ḍāti-al-waqti kāna al-<irāqu huwa al-qāri'u al-nahimu..
[The publishing houses in Egypt and Lebanon were flooded with new manuscripts that were waiting to take their due place in the Arabic library. At the same time, Iraq was the insatiable reader ...]
(J, 24/5/83, X100, 19103-7)

6.9.4 Temporal Span

6.9.4.1 Repertoire

The following list comprises connectives that signal the temporal relation of time span:

A. English

Simple: henceforward, since, till, until

Compound: since then, until then.

B. Arabic

Simple: ḥattā, mundu, 'illā, mundu'iḏin

Compound: 'ilā 'an, 'illā wa, mundu ḏālika al-waqti

6.9.4.2 Textual Functioning

The basic function of the connectives of this category is to introduce a proposition that specifies the time span during which the other proposition is valid, or the time limit beyond which it is not valid. The connective in this case sets a time dimension with a hypothetical point representing the current valid state. This point represents the time of primary concern to the text producer within the given textual context (that is, the point has an orientation to the text producer's "now"). The connective then manipulates the conceptual contents of the proposition it introduces (in the case of since, till, until in English, "'ilā 'an", "ḥattā", "munḏu", in Arabic), or refers to (in the case of since then, or until then) so as to set up a span along the time dimension and specify a limit point. Within this span the conceptual content of the first proposition is projected and assessed.

It should be mentioned that the connectives until in English and "'ilā 'an" and "ḥattā" in Arabic occasionally carry a shade of causal meaning. In these instances the antecedent proposition establishes a cause and the subsequent (the one that is introduced by the connective) specifies the limit beyond which there is a

consequence. This causal meaning heightens the relation of temporal span and renders it more forceful. The element of causality is evident when the proposition that expresses the time span follows the other proposition. It is also more noticeable and more forceful in Arabic than in English. Indeed the causal meaning of the connective "ḥattā" can be reinforced to the extent that it starts to signal causal rather than temporal relation. Examples from the corpora:

[6.142] ... highly competitive international industries such as steel, cars, textiles and chemicals will continue to be depressed in the US until the budget deficit and in turn the trade deficit are trimmed.
(G, 19/11/82, X3, 427-31)

[6.143] ..yamuttu fī marāḥili tanfīḍi-hā ḥattā yatajāwaza-hā al-zamanu..
[They (planners) prolong the stages of implementation until their plans get obsolete...]
(Nb, 21/5/83, X12, 2045-7)

6.9.4.3 Some Textual Patterns

There are two types of time span depending on its location in relation to the text producer's "now" point. A forward span occurs after that point, i.e. it represents future in relation to the "now" point, while a backward span occurs before it, i.e. it represents a past span. The forward span is signalled by the connectives until, until then, henceforward in English, and by "ḥattā", "'ilā 'an", in Arabic. Examples are given above in [6.142-3]. Backward span is signalled by since, since then in English, and by "mundu", "mundu'iḍin", "mundu ḍālika al-waqti" in Arabic. The following are examples from the corpora:

[6.144] In the four weeks since the negotiations resumed in Geneva the Russians have hardened their insistence on having these weapons included in the European nuclear balance.
(O, 12/6/83, X82, 13768-71)

[6.145] laqad 'arada l-i tawfiqu allahi, mundu bada'-tu hayat-i al-<aqliyyati al-muntijati, 'an 'aqa<a <ala tariqin min turuqi al-tafkiri al-falsafiyyi.. [God has willed since I started my intellectually productive life to follow a certain philosophical approach ...]
(Hr, 24/4/83, X53, 9686-88)

6.9.5 Temporal Positioning

6.9.5.1 Repertoire

The following lists comprise connectives that signal temporal positioning:

A. English

Simple: as, later, once, previously, then, thereafter, when.

Compound: as soon as, at a moment (when), at a time when, at [just] that point, at one time, at precisely the moment, at such moment, at that moment, at this [that] stage, at that time, at the time, at this point, at those moments, by then, by the time, during a time, from the moment, in such [these] circumstances, in such a case, in such an event, in that event, in the circumstances, in the days when, in the event, in the period when, in this context, in those days, on that [another] occasion, on the same day, on the same occasion, on these occasions, on this occasion, the moment, this time.

B. Arabic

Simple: <indamā, hīna, hīnamā, 'iḏā, <inda'iḏin, yawma, 'ānaḏāka, ḥattā, ḥayṯu, ḥīna'iḏin, lammā, hunā, ḥīnaḏāka, ba<damā, 'ayyāma, laḥẓata, sā<ata'iḏin, waqta, yawma'iḏin.

Compound: fī al-waqti alladī, fī ḥīni, fī waqtin, <inda-hā, fī ḏālika al-waqti, fī hāḏihi al-marḥalati, hāḏihi al-marrata, 'iḏ ḏāka, fī ḥīni-hi, fī hāḏā al-waqti alladī, fī haḏihi al-ḥālati, mā 'an, bi-al-waqti, dāxila hāḏihi-al-fatrati, fī al-laḥẓati allatī, fī al-marrati, fī al-yawmi nafsi-hi, fī ḥālati mā, fī bad'i, fī waqtin sābiqin, fī hāḏā al-ẓarfi, fī hāḏā al-wāqi<i, fī hāḏihi al-fatrati, fī hāḏihi al-laḥẓati, fī ḏālika al-zamāni, min al-bidāyati, min al-laḥẓati allatī, min hāḏihi al-marḥalati

Correlate: ma [lan, la, lam]...ḥattā

6.9.5.2 Textual Functioning

Connectives of this category relate two knowledge configurations by specifying a point on the time dimension of one and positioning the other onto it. Generally, this function helps keep the time dimension of the textual world current by constantly updating the temporal positions of the various propositions, a factor that will eliminate possible ambiguity of time reference. In information terms, connectives can re-activate a previous knowledge configuration by setting a time location for the next one. Alternatively they can pro-activate a subsequent knowledge configuration by relating the current one to it. This activation is

automatic and is meant to reduce processing load by cutting on backtracking, by reducing waste caused by the amount of search made to check temporal relevance.

The time location may be narrow or wide according to the conceptual space of the connective itself. For instance, at that moment pinpoints a narrower position that on these occasions on the time dimension. However, this consideration has to be accepted as relative and it depends on the transparency of the time dimension itself as expressed in the relevant proposition, or in the text as a whole.

Some connectives of temporal positioning may at the same time express sequence, and hence they lie on the border of two subcategories. For example, the connectives when and as soon as, in English and the Arabic connectives "<indamā" and "ḥālamā" may signal duality or even multiplicity of time relation (see examples in the next section).

Another aspect of textual functioning is that some temporal connectives express causality as well (see examples below). In these cases, the connective indicates that a particular propositional content (e.g. event or state) is not only positioned temporally along the time dimension of another proposition, but is dependent causally on it. This duality of function strengthens the cohesive bond between the two propositions.

6.9.5.3 Some Textual Patterning

The types of functions discussed above suggest a number of patterns that have been identified in the two corpora, though with different distribution:

1. The temporal positioning "proper": the connective indicates time reference only. Most of the connectives having this pattern contain an anaphoric reference that consolidate the signalling of the function. Generally, the positioning can be made in two ways.

a) there may be an explicit time reference, as in the expression "to the 1982 budget" in [6.146], and the next proposition is related directly to it.

[6.146] To understand why, you have to look back to the 1982 budget. At that time Sir Geoffrey predicted that public borrowing in 1982-83 would be 9.5bn.
(STi, 24/4/83, X237, 37161-4)

[6.147] ..wajadat al-'idāratu al-'amīrkiyyatu fī ḏālika-al-waqti 'anna rafḍa 'isrā'ila li-al-'insiḥābi sawfa yu<arriḍu ṭumūḥa-ha li-al-xaṭari..
[... the American administration at that time found that Israel's refusal to withdraw (from Egypt) would endanger America's ambition ...]
(Nb, 5/4/83, X4, 406-9)

In [6.147] the connective "fī ḏālika al-waqti" [at that time] refers to "December 1956" mentioned in a previous paragraph.

b) There is no explicit time reference. The whole conceptual content of a proposition is treated as a point where the other proposition is temporally positioned. This is a more common pattern than the previous one. Examples:

[6.148] A 75-year old from the East caught the tragic dimension of what has happened when he said "Everybody was a socialist when I was a kid; only because they didn't know it, they didn't know when they were losing it."
(G, 17/1/83, X13, 1932-7)

[6.149] ..<indamā 'intahā wujūdu-hu kāna qad taraka al-musta<marāti nahban li-al-'inqisāmāti al-ṭā'ifiyyati..
[... when its presence [of Western occupation of Arab countries] was ended, it had left the occupied countries in a turmoil of sectarian division ...]
(Sh, 18/3/83, X120, 23228-30)

2. Time positioning is often conflated with sequencing. However, sequencing is not the focus; rather it is a means of identifying where on the time dimension each proposition is positioned. Examples:

[6.150] ... when a baby was presented to be held he [Michael Foot] stopped and held it. The baby hated him.
(O, 29/5/83, X74, 12263-5)

[6.151] ḥīna ṣadara qarāru al-'umami al-muttaḥidati bi-taqsīmi filasṭīna <āma 1947 rafaḍa al-<ālamu al-<arabiyyu hāḏā al-qarāra.
[When a resolution was passed in the United Nations in 1947 for the division of Palestine, it was rejected by the Arab world.]
(Ar, 5/11/82, X152, 29011-13)

3. Temporality is often conflated with causality; and connectives such as as, once, then, when in English and most Arabic temporal connectives are capable of combining both functions. However the temporal meaning is normally dominant.

[6.152] Accidents happen when people are negligent.
(G, 28/3/83, X20, 3147)

[6.153] nazala al-xabaru <alā ra's-i ka-al-ṣā<iqati ḥīna 'ablaga-nī ṣadīqun 'anna al-mušīra muḥammad <abd alganī aljimasī qad 'asbaḥa mudīran li-'iḥdā šarikāti al-qiṭā<i al-xāṣṣi al-<āmilati fi ṣinā<ati al-ruxāmi..
[The news came like a thunderbolt on me when I was informed by a friend that Field Marshal Al-Jimasi is now a director of a private sector company for the manufacture of marble ...!]
(Nb, 21/5/83, X13, 2086-89)

4. The connectives at a time when and its Arabic counterpart "fī waqtin" may combine an adversative meaning with temporality. The propositions after these connectives not only express a temporal positioning for the antecedent, they also indicate unfavourable or contrary circumstances.

[6.154] ... the Referendum will cost up to 1 million at a time when Ireland is desperately short of money.
(O, 23/1/83, X49, 7756-7)

[6.155] kullu hāḏā yaḥṣulu fī al-waqti alladī yujābihu fī-hi al-waṭanu al-<arabiyyu hajmatan šarisatan min al-'imbiryāliyyati wa al-ṣahyūniyyati..
[All that happens at a time when the Arab homeland is exposed to a ferocious campaign [attack] carried out by imperialism and Zionism ...]
(Ar, 3/6/83, X203, 34992-4)

A final note is on the Arabic correlate with "hatta" as its core, the nearest equivalent in English being "no sooner ... than". The antecedent is in the negative using such particles as "mā", "lā", "lan", or "lam". The temporal relation is intensified and the subsequent is placed in a climatic focus. The correlate normally combines a causal as well as a sequential element of meaning, which together produce a sense of textual heaviness and salience. The following excerpt is an example:

[6.156] ..lā yaxtafī dawrun ḥattā yabda'a dawrun..
[No sooner a phase [of the Zionist aggression] is ended than another starts.]
(Sh, 1/3/83, X114, 21981-2)

6.9.6 Temporal Circumstance

6.9.6.1 Repertoire

The connective that signals this type of relation is "wa" in Arabic.

6.9.6.2 Textual Functioning

In signalling this relation, the connective "wa" has, broadly, a meaning of temporal simultaneity, and hence one can argue that it is a variant pattern of that textual function. However, the fact that this pattern has interesting structural as well as functional

features of its own makes it one of the peculiarities of Arabic, and hence worth occupying a subcategory of its own. (For the various structural patterning of the circumstantial clause in Arabic, see Cantarino Vol. III pp. 266-279).

Specifically, "wa" in this function introduces a proposition that specifies the attendant context or circumstance to the current main proposition as a whole or to one of its concepts. However, the relation between the attendant circumstance and the main proposition is variable and may be classified into five broad types:

1. Purely temporal: This function is similar to that of temporal simultaneity. The relation between the circumstance to the main proposition is that of two events or courses of action taking place simultaneously.

[6.157] majmū<atun ḍaxmatun min hāḏā al-naw<i min al-'as'ilati ṭaraḥtu-hā <alā nafs-ī wa 'anā 'ufakkiru bi-hāḏā al-<amali al-kabīri jiddan.
[I asked myself many such questions as I was thinking of this great deed.]
(Sh, 29/11/82, X106, 20384-6)

2. Explanatory: The attendant circumstance establishes an explanation. For example:

[6.158] ma<a al-naksati wulidat 'ibnat-ī. 'imtazajat farḥat-ī bi-mīladi-hā bi-al-'alami ya<taṣiru 'a<māq-ī wa naḥnu na<īšu wāqi<a al-hazīmati.
[My daughter was born during the [June 1967] crisis. My joy was mixed at her birth with bitterness that pinched the depth of my soul as we lived the reality of defeat].
(Nb, 4/4/83, X1, 23-5)

3. Antithetic: The attendant circumstance implies something unexpected in the simultaneous validity of the two propositions.

[6.159] wa al-su'ālu: limāḏā yakḏib-ūna wa hum ya<rif-ūna qabla gayri-him 'anna-hum mā yazāl-ūna... yajurr-ūna 'aḍyāla al-xaybati wa al-xasārati wa al-xuḏlāni wa bi-šaklin mutawāṣilin?
[And the question: why do they lie when they know before anyone else that they still continuously suffer disappointment, loss and defeat]
(J, 7/5/83, X98, 18837-42)

4. Descriptive: The attendant circumstance offers a specification of the manner in which the content of the main proposition is formulated, as in [6.160], or a description of one of its concepts [6.161].

[6.160] wa 'ajāba-nī wa nabratu al-jiddiyyati lam taxtafī min sawti-hi: na<am; qāmūsun yaḍummu kulla al-muṣṭalaḥāti allatī daxalat 'ilā "mu<jami" al-siyāsaṭi al-<arabiyyati fī al-sanawāti al-ṯalaṯīna al-'axirati.
[And he replied with the serious tone still obvious in his voice: yes a dictionary that comprises all the terms that have entered the Arab political lexicon during the last thirty years.]
(Nb, 5/4/83, X8, 938-42)

[6.161] ṣawtu al-'ālati wa hiya tadūru yuṭribu-nī wa ka'anna-hā tugannī.
[The sound of a machine as it is working enchants me as though it were singing.]
(Sh, 29/11/82, X105, 20189-90)

5. Orientative: The attendant circumstance offers the angle or perspective from which the main proposition is to be viewed.

[6.162] 'inna hāḏā al-mawqifa al-ḥāqida al-taxrībiyya al-la'īma li-hāḏayni al-niẓāmayni lam ya<ud wa al-ḥālatu hāḏihi mujarrada mawqifin xiyāniyyin maš̆būhin, 'innama huwa mawqifun ta'āmuriyyun <udwāniyyun ṣarīḥun..
[This mean and spiteful act of sabotage of these two regimes is no longer, under such circumstances, a mere dubious political attitude; it is a plain act of conspiracy.]
(Th, 19/3/83, X70, 13011-4)

6.9.7 Time Frequency

6.9.7.1 Repertoire

There is a short list of temporal connectives that have the function of indicating time frequency.

A. English

Simple: whenever

Compound: every time

B. Arabic

Simple: kullamā, mā

Compound: mā dāma

Correlate: kullamā...kullamā

6.9.7.2 Textual Functioning

Connectives of this category express a conditional and temporal meaning. The conditional function is based on the dependency of the content configuration of the subsequent proposition on the content of the antecedent. That is to say, if the antecedent (conditioning) proposition is validated, the subsequent (conditioned) one will also be so; if it is not, the consequent is equally invalidated. This explains why Arab grammarians consider "kullamā" a conditional particle. However, this conditional meaning is often weakened or neutralised to give a more temporal meaning of repetitive or habitual contingency. Repetitiveness is conflated with temporal positioning; that is to say, in addition to the expression of recurrency, the propositional content of the antecedent expresses a temporal point where the content of the consequent takes place. This is indicated by the morphological composition of whenever and

"kullamā". As for the Arabic connectives "mā" or "mā dāma", temporal contingency is based on duration rather than positioning, the interpretation being a temporal "as long as". The following are examples on the textual functioning of this connective.

[6.163] ... many men appear to be immeasurably flattered, whenever an intelligent woman pays them careful attention.
(G, 8/6/83, X42, 6708-10)

[6.164] 'inna-nī 'aḥzanu kullamā waqa<at <ayn-ī <alā qaṣā'ida kaṯīratin 'aqra'u-hā fī hāḏihi al-'ayyāmi wa lā 'afhamu mā yurīdu al-šā<iru 'an yaqūla-hu.
[I feel sad whenever I see poems that I read but can't understand.]
(Sh, 9/6/83, X135, 26243-5)

6.10 Spatial

6.10.1 Repertoire

The following list of connectives signal spatial relations.

A. English

Simple: elsewhere, therein, where, whereby, wherever,

Compound: at a point (where), at just that point, at one point, to the point (where).

B. Arabic

Simple: ḥayṯu, ḥayṯuma

Compound: fī haḏa al-majāli, min-ḥayṯu

6.10.2 Textual Functioning

Spatial connectives are referential in essence. Their basic function is to provide spatial orientation whereby the subsequent proposition is related to the current one. Usually the relation is established in terms of spatial positioning, that is, the

propositional content of the consequent is related to a particular point on the space dimension, and the connective functions as a location indicator. Spatial connectives are therefore limited in function and lack the varied functionality of the temporal connectives in relation to the time dimension. A more elaborate space definition is usually achieved by adverbials, by adjectives and even by verbs. These space definers either specify relationships of separately positioned objects, locate a positional juncture or indicate a direction (see the discussion of place adjuncts and place relators in Quirk et al. 1972, 1985). The following excerpts are examples from the corpora:

[6.165] ... in both the safe Labour and safe Conservative seats, Labour did worse where the Alliance did best ...
(STel, 12/6/83, X143, 23952-54)

[6.166] ..yasīra ḥaytu ya'muru-hu <alā 'asāsi 'anna ḏālika wājibun (dīniyyun)..
[... they [the Iranian people] go where [the Iranian regime] command them, on the basis that it is a "religious" duty ...]
(J, 17,3, 83, X89, 17213-4)

A variant of the function is where the connective sets up a relationship, not in terms or real space, but in terms of figurative location. In this case an imaginary place dimension is constructed within the antecedent, and then a position is pinpointed, to which the subsequent proposition is related. Examine the use of where and "ḥaytu" in the following two examples respectively.

[6.167] Where questions of conscience are involved, it is never right for the majority to ride roughshod over the feeling of the minority.
(STel, 26/6/83, X148, 24791-4)

[6.168] ..yuṣirru fī-hi <alā <adami al-naẓari bi-al-waḍ<i al-'iqtiṣādiyyi al-sayyi'i, wa yatagaḍā <an

mu<addalāti al-taḍaxxumi fī al-'iqtiṣādi ḥaytu al-'as<āru murtafi<atun wa al-mustawā al-ma<āšiyyu sayyi'un li-al-gāyati..
[... it [the Israeli government] insists on avoiding consideration of the deteriorating economical situation and on ignoring the (high) inflation rates where prices are high and the standard of living is extremely bad ...]
(J, 3/5/83, X97, 18753-7)

It is interesting to note that while English connectives of place are distributed fairly evenly in texts from various newspapers, a substantial number of the occurrences of Arabic place connectives, particularly "ḥaytu", is concentrated in the Iraqi newspapers, in texts written by Iraqi writers. A detailed account of this finding goes beyond the scope of this work, as it involves possible regional variations within Modern Standard Arabic.

6.11 Causal

6.11.1 General Comments on Textual Functioning

A causal relation represents nexus of two (sets of) knowledge configurations where one expresses a cause/condition while the other expresses a related effect/consequence. To understand this relation we must examine some of its main features.

1. The two (sets of) propositions create in their nexus a causal field where they are seen as "changes" or "differences", either real, potential or hypothetical. This causal field is operative within a particular world, i.e. the text world or knowledge world, and is, therefore, constrained to a large extent by factors, or "circumstances", that are active in this world.

2. The propositions in the causal field exhibit some kind of dependency: the determinacy of one statement is contingent upon

access to the other. In other words, to say that statement A causes statement B means that B represents a change that is dependent for its occurrence on A.

3. Dependency in the causal field is explainable in the notion of priority ("causal priority" in Mackie's 1980 view). The core of this notion is that cause has a prior existence to its potential effects. Priority is further distinguished by two facets that act as constraints in a causal field. The first one is sufficiency. Statement A is sufficient for statement B in the text world and is causally prior to B provided that if A is placed into the text world and the world runs from there, B will occur. The second facet is necessity. Statement A is necessary in the text world for statement B and is causally prior to B provided if A were removed from the text world, and the world is allowed to run on from there, B would not occur.

4. It follows that the causal relation is asymmetrical in the sense that it does not allow commutativity. A connection where proposition A is the cause for proposition B does not entail that B is the cause for A. If one is to challenge this by contriving examples where commutativity is seemingly permissible, then one is creating different causal fields with different priority and direction of dependency, and therefore two different causal connections.

5. The causal relation can be implicit; it is inferable from content within the propositions making up the causal field. In this case processing ease is decreased to a variable extent, depending on

the scale of informativity of the two propositions (which is usually determined by the size of processing resources expended upon the input, cf. Beaugrande 1980). Implicit relations are not the focus of this study. Our interest lies in those causal relations that are made explicit through the use of causal connectives. This provision of overt surface signalling can contribute to efficiency of processing (as long as their use is not unduly frequent, a point that brings up the question of the probabilities of surface signalling of various relations). It not only marks off the presence of causality, but indicates its type, direction and range.

6. Causal connectives can be categorised into various types according to the way they represent the causal field, and the semantic interpretation they impose in order to understand it. In the corpora we have distinguished five types of representation: a) the next stretch of text (proposition or set of propositions) represents the cause of the current stretch of text (the antecedent or subsequent); b) the next stretch of text represents the effect or consequence in relation to the current stretch of text or indicate an inference or conclusion drawn from it; c) the causal field represents a conditional relation; d) the connectives indicate degree, scale or extent of a cause and its impact on the effect; e) the relation involves expression of purpose. These will be discussed in the next five sections with the connective repertoire and textual function of each. It should be noted that types (a) and (b) mentioned above can be combined into one type. Since propositions connected via causality are non-commutative, the main difference between (a) and (b) is that of direction of the causal

dependency, as indicated in the following:

(a) effect/consequence/inference ⟶ cause/reason

(b) cause/reason ⟶ effect/consequence/inference

Further, one may be justified in subcategorising (b) into two types of function: one focuses on result/effect/consequence while the other concerns inference and conclusion. The division is made on pragmatic basis, mainly because it facilitates the contrastive account to be conducted later.

6.11.2 Cause/Reason

6.11.2.1 Repertoire

The following list comprises connectives that signal cause or reason.

A. English

Simple: as, because, for, lest, now (that), since

Compound: as long as, for the same reason (that), in that, on that account, on the assumption (that), on the calculation (that), on the ground(s) (that), so long as, insofar as.

Correlate: as ... so.

B. Arabic

Simple: fa, li'anna, ḥayṯu, wa, 'iḏ, ṭālamā, lammā, ḥasbamā.

Compound: ḏālika 'anna, mā dāma, bi-'i<tibāri, bi-ḥujjati, wa ḏālika, <alā 'asāsi, 'istinādan 'ilā, bimā 'anna, wafqan li, min hāḏā al-munṭalaqi, wa ḏālika bi, ṭibqan li, bi-sababi, <alā 'i<tibāri, 'amā wa,

'intilāqan min, li-mā, naẓaran 'ilā, wafqa hāḏā, bi-ḥukmi, li-hāḏihi al-'asbābi, min ḥayṯu, min haḏihi al-zāwiyati, 'inṭilāqan mimmā, 'istidlālan mimmā, binā'an <alā, <alay-hi, <alā ḍaw'i, li-'asbābin, li-nafsi al-'asbābi allatī, min munṭalaqi

6.11.2.2 Textual Functioning

The connectives of this category signal the cause proposition in the causal field, and this indicates that the other proposition (or set of propositions) is dependent on it. Put in different terms, the connective labels a conceptual link between nodes and indicates the traversal direction. Occasionally the link is between minor networks each with its own conceptual centres. The connective thus helps the processor to traverse the links without unnecessary checking or excessive pattern matching.

The relation of cause can be classified into two main types:

a. Cause "proper": The connective introduces the cause that has led to the event, state or course of action expressed in the effect proposition (or set of propositions). Examples:

[6.169] ... medical students and hospital interns are worried about the future because the system produces far too many doctors ...
(STi, 29/5/83, X247, 38809-11)

[6.170] fi kulli maṭārin lā yusmaḥu la-hu bi-al-nuzūli li'anna "waṯīqata al-safari" al-lubnāniyyata allatī yahmilu-hā qad 'intahat muddatu sarayāni-hā.
[In every airport he is not granted entry visa because his Lebanese "travel document" has expired.]
(Sh, 9/6/83, X134, 26085-8)

b. Explanation: In this function the connective introduces an explanation or justification for an event, state or course of action. This is a frequent function for the Arabic connectives of cause and has a fundamental role in organising the Arabic paragraph, as will be discussed later (cf. Chapter 9). The function is most frequently signalled by for in English and "fa" in Arabic.

[6.171] The liberation of the Third World turns out to be a struggle that doesn't threaten us at all; for it is an indivisible part of our own.
(G, 17/1/83, X13, 2128-30)

[6.172] maxawifu maṣdaru-hā ẓulumātu al-jahli. fa naḥnu la na<rifu ṭabī<ata tilka al-'ajsāmi..
[Fears the origin of which is ignorance. For we do not know the nature of those objects ...]
(Hr, 25/1/83, X33, 6455-6)

6.11.3 Result/Inference

6.11.3.1 Repertoire

The following list comprises connectives that signal effect or result.

A. English

Simple: accordingly, and, consequently, so, then, thereby, hence, therefore, thus.

Compound: so that, as a consequence, as a result, for a different reason, for some reason, for that [this] reason, for the same reasons, for these reasons, for whatever reason, in consequence, in turn, in the light of (those), on this basis, with the result (that).

B. Arabic

Simple: wa, fa, hākaḏā, 'iḏan, mimmā, liḏā, li.

Compound: li-ḏālika, bi-al-tālī, min hunā, bi-ḥayṯu, bi-ḏālika, li-hāḏā, wa min ṯamma, al-'amru alladī, bi-mā, ma<nā hāḏā, bi-hāḏā, ma<nā ḏālika, li-hāḏā al-sababi, wa ḏālika, <alā hāḏā al-'asāsi, min natījati, ma<nā min ḏālika, natījatan li, wafqan li, bi-al-natījati, bi-sababi, <alā ḍaw'i, fī ḍaw'i, min 'ajli, binā'an <alā ḏālika, binā'an <alay-hi, bi-faḍli, <alā hāḏā al-mi<yāri, ka-ḥasīlatin li-mā, ka-natījatin, taḥta hāḏihi al-ḥujjati, wa ḏālika bi.

6.11.3.2 Textual Functioning

Connectives of this category introduce the consequence in a causal field, where the cause is expressed in the conceptual content of another proposition (or a sequence of propositions). Connectives of both categories cause and effect may convey temporal meaning. This is because the notion of priority that we have mentioned above often imposes an ordering in time, whereby the content of the cause proposition is temporally prior to that of the effect.

Examination of the behaviour of connectives of this category suggests that the function of signalling result/inference fluctuates between strictness in the expression of causality in which the connective has the meaning of "because of that" (examples [6.173-4] below) to vagueness of causal reference where the connective may have a partly summative, partly conclusive meaning, similar to "it follows" (examples [6.175-6]):

[6.173] ... production will require that labour and capital be transferred from some other sector of the economy. As a result, production in this sector will fall.
(Ti, 22/11/82, 27047-50)

[6.174] hal hiya dawlatun musta<miratun wa taxḍa<u al-<adīdu min al-duwali li-sayṭarati-hā wa maṣāliḥi-hā, wa tataḥammalu natījatan-li-ḏālika mas'uliyyata ri<āyati-hā?
[Is it an imperialist state, dominating a number of other states and for this reason has to look after them?]
(Sh, 29/11/82, X106, 20329-32)

[6.175] The romance was a grand, joyous game, which Diana won. She just didn't reckon with having to play it forever, all the time. So, shy Di became, first the diet-obsessed anorexic Di, then hen-pecking, scolding Di, and now, the little madam or fiend, or spoilt brat, according to whom you read.
(STi, 23/1/83, X233, 36667-73)

[6.176] ..tusabbibu al-'aḍrāra bi-al-maṣāliḥi al-barīṭāniyyati, mimmā sa-yu'addī 'ilā xasārati al-barīṭāniyyīna lī-'ulūfi al-waẓā'ifi wa malāyīni al-junayhāti..
[... she [Mrs. Thatcher] is inflicting damage on the British [trade] interest, which (as a result) may cause the British to lose thousands of jobs and million of pounds in foreign trade.]
(Ar, 27/1/83, X181, 31670-3)

This fluctuation of the function varies from English to Arabic as will be discussed later (see 9.2.9.2.2 in Vol. 3).

Related to this point is the observation that some connectives, particularly and in English and "wa" and "fa" in Arabic, conflate causality with additivity and, often, with temporality as well. This semantic blend renders the connectives less strict themselves in expressing causality, which explains why linguists such as van Dijk (1977a) call "and" in this context a "neutral" connective. However, it should be noted that the expression of causality in these cases is often signalled lexically within the consequent, or

supported by its conceptual content.

Another textual function of the connectives of this group is the expression of an inference or a conclusion extracted from the antecedent proposition. The connectives that signal this function in English are hence, then, therefore and thus. In Arabic it is expressed via the connective "hakada" and a number of compound connectives that are made up of a preposition (li, bi, min) and one of a small number of referential items that refer to the antecedent proposition(s). These connectives are "li-hāḏā", "li-ḏālika" [therefore, thus, hence], "bi-hāḏā", "bi-ḏālika" [by this, that], "min hunā" [hence], "wa min ṯamma" [then] and "wa hunā" [by now, thus]. Examples:

[6.177] The key principle of GATT [the General Agreement on Tariffs and Trade] is that trade arrangements between countries should be non-discriminatory. If, for example, a GATT member has a 10 percent tariff on the import of cars from one another member, the same 10 percent tariff should apply to the import of cars from every other member. Thus the system is multilateral, not bilateral.
(Ti, 22/11/82, X165, 26911-17)

[6.178] ..ma zālat jāmi<ātu-nā tudarrisu al-qissata al-qaṣīrata <alā 'anna-hā "tsīkūf". wa mā zāla al-šī<ru al 'inkilīziyyu fī naẓari ba<ḍi jāmi<āti-nā yatamaṯṯalu fī "'ilyūt". wa li-hāḏā fa 'inna hāḏihi al-jāmi<ātu tu<ānī min ḥālati 'infiṣālin <an al-tayyārāti al-ṯaqāfiyyati fī al-<ālami.
[... our universities still teach the short story as represented by Chekhov. Further, the English poetry is still, in the opinion of some [scholars], represented by "Elliot". Thus these universities suffer from isolation from the [current] international cultural trends.]
(Sh, 1/3/83, X115, 22226-32)

6.11.4 Magnitude/Degree

6.11.4.1 Repertoire

The following connectives signal a degree/result relation:

A. English

Compound: to the extent (that), to the point (where, that)

Correlate: so ... that, such ... that.

B. Arabic

Simple: hatta

Compound: 'ilā ḥaddi [ḥaddin], 'ilā (hāḏā) al-ḥaddi, li-darajatin, 'ilā darajati, bi-qadri mā.

6.11.4.2 Textual Functioning

Connectives of this category signal a relation whereby the consequent expresses a result of the intensification of all or part of the conceptual content of the antecedent. There is some overlap between this function and the function of the comparative connectives of degree. Both aim to bring a certain concept (or a set of concepts) into focus; but then they diverge in the way this focus is manipulated: the comparative connectives match the element in focus with the content of the consequent in terms of equivalence or nonequivalence, while the causal connectives treats the element (or set of elements) in focus in terms of insufficiency or excess and relate it to an actual or potential consequence.

6.11.4.3 Textual Patterns

The common pattern for the connectives of this category is to perform a double function: they carry out the conceptual intensification mentioned above and, simultaneously, identify its consequence. The two functions are fused together. Examples:

[6.179] The Press is taking the lid off some amazing local police scandals. The police in Chelyabinsk region was so rotten that the party has had to draft 1800 trusted factory workers into the force.
(O, 8/5/83, X59, 9576-9)

[6.180] 'inna-ha 'arqamun mudhilatun li-darajatin lam tattasi< la-ha xanatu al-makinati al-hasibati
[These are stunning figures to the extent that the digits of a calculator will not take them.]
(Sh, 29/11/82, X106, 20249-50)

In another less common pattern, which is peculiar to Arabic, the connective performs one function. It does not express conceptual intensification; this is already implicit in the content of the current proposition. It only relates the consequence to the already intensified antecedent. This is the pattern in which the connective "hatta" is normally used. For example:

[6.181] ..'irtafa<at mabi<ātu al-silāḥi al-faransiyyi li-al-su<ūdiyyati ḥattā ṣārat fī al-martabati al-ṯāniyati ba<da al-'asliḥati al-'amīrkiyyati.
[... French arms deals to Saudi Arabia have expanded to such a degree that they have been next only to American arms deals.]
(Ar, 27/1/83, X141, 31958-61)

6.11.5 Purpose

6.11.5.1 Repertoire

The following connectives have the function of signalling the causal relation of purpose.

A. English

Simple: lest

Compound: in case

Correlate: so that

B. Arabic

Simple: li, ḥattā, likay, kay, li'allā, kaymā, likaymā.

Compound: min 'ajli, li-hāḏā al-garaḍi, bi-hadafi 'an, fī sabīli ḏālika, fī sabīli hāḏihi al-gāyati.

6.11.5.2 Textual Functioning

Connectives of this category signal a type of contingency between two (or more) propositions, whereby, in the textual world, proposition B makes a justification or a specification of the aim or purpose of proposition A. As with other types of concatenation based on causality, this relation helps organise text by integrating information blocks: a knowledge configuration is supported by another and the type of relatedness and dependency that obtain keeps it integrated in the textual world.

[6.182] ... Conservative, Labour and SDP politicians all covered up their previous policies, so that revenge for the national humiliation could be meted out in the colours of principle.
(G, 28/3/83, X19, 2663-7)

[6.183] wa al-lubnāniyyu waṣala 'ilā marḥalatin 'iqtiṣādiyyatin mut<abatin wa yaḥtāju 'ilā al-'amni ḥattā yaštagila fī rizqi-hi.
[And the Lebanese have reached a bad economical stage and so needs security so that they may work for their bread.]
(Sh, 20/4/83, X124, 23889-91)

6.11.6 Condition

6.11.6.1 Repertoire

The following connectives signal a conditional relation:

A. English

Simple: allowing, assuming, else, given, if, or, otherwise provided, providing, supposing, unless.

Correlate: if ... then.

B. Arabic

Simple: 'iḏā, law, 'in, lawlā, man, la'in.

Compound: 'illā 'iḏā, wa 'illā, <alā 'an, mā lam, <alā 'allā, 'illā wa, bi-šarṭi, <alā šarti, <alā 'iftirāḍi.

Correlate: lā...'illā, lā...'illā wa.

6.11.6.2 Textual Functioning

Conditional connectives signal an interdependency of two knowledge configurations, whereby the presence of B in the textual world is dependent on the existence of A. That is B does not follow unless the condition stated in A also holds. However the relation leaves unresolved whether proposition A exists in the textual world and therefore it leaves unresolved whether proposition B is fulfilled. This openness as to the factuality of the condition is one of the major characteristics of this relation.

This relation can be of several types:

1. Hypotheticality: This is conditionality 'proper'. The connective indicates that the truth of proposition B is a consequence of the condition expressed in proposition A.

[6.184] If Britain tries to be nasty to South Korea by imposing quotas on its steel and ships, South Korea can answer back by refusing to buy British power-station equipment and textile-making machinery.
(Ti, 22/11/82, X165, 27009-13)

[6.185] fa 'in lam nabda' al-<amala fawran fa 'inna lā <uḏra la-nā..
[So if we do not start acting immediately then there is no excuse for us ...]
(Sh, 25/6/83, X147, 28493-5)

2. Contrafactuality: This is a type of hypotheticality which conveys the text producer's belief that the condition in proposition A cannot be or was not fulfilled in the current textual world. This

is because it runs contrary to expectations, assumptions or facts. It follows that the content of proposition B is probably or certainly false. For instance, the following example:

[6.186] If the two major parties in our political system were both free of socialism and union power, that would be the happiest political achievement since the war.
(DTel, 8/6/83, X121, 20201-5)

is an expression of the function of hypotheticality where proposition B is dependent upon proposition A. However, since proposition A runs contrary to the factuality of the textual world, there is an implied negative meaning, i.e. that "the two major parties in our political system are not free of socialism and union power", which, in turn, implies the negation of proposition B, or the non-existence of its content. In English, this function is marked on the surface level by backshifting the verbs in the clauses. In Arabic, however, backshifting is rarely indicative of contrafactuality. Instead, a specific connective, "law", may be used. Alternatively, the propositional content of the clauses themselves will give sufficient indication. For example:

[6.187] law kāna la-hā ṣawtun la-jā'a xalīṭan min 'anašīdi al-'amali wa 'ajrāsi al-ḥuzni.
[if [the past years of my life] had a voice of their own, it would be a mixture of songs of hope and bell tolls of sadness.]
(Nb, 4/4/83, X2, 104-6)

3. Rhetorical condition: In this subcategory we include all instances of conditionality that, though they have the surface structure of hypotheticality, actually convey or shape the text producer's views or assertions. In English, the rhetorical condition can be of several types, some of which are:

a. The antecedent (Proposition A, which is introduced by the connective) may function as a framework or angle from which the consequent is viewed.

[6.188] If this is so, we have every right to request that before permits to work are issued to foreign labour, British seafarers should first be offered the jobs.
(DTel, 20/4/83, X107, 18013-16)

[6.189] Bilateralism may be all very well if one looks at it from the point of view of a single country pursuing it with enlightened self-interest.
(Ti, 22/11/82, X165, 26898-900)

Sometimes the assertion may take the form of a conclusion:

[6.190] If he is right, poor Britain, uniquely, have got the worst of the 70's price boom, and the 80's slump. (STi, 23/1/83, X230, 36372-4)

[6.191] It was a gloomy but persuasive view of Britain's post-war development. And if it wascorrect, we must conclude that the heart and soul of Thatcherism and its remarkable success is the emergence of a new mood of national pessimism.
(G, 6/6/83, X34, 5522-6)

b. The antecedent presents a view that is, within the text producer's sequence of arguments, considered patently absurd in order to prove that the statement in the consequence is not, or cannot be, true, or in order to pose a rhetorical question.

[6.192] ... if the Tories are really so "extreme" as the Alliance claims, why have they been for months so far and so persistently ahead in the opinion polls ...
(O, 29/5/83, X76, 12701-4)

c. The antecedent may express a denial of a certain view or conclusion in order to allow the expression of a more desirable one or the expression of a rhetorical question.

[6.193] The 1945 Conservative manifesto is Churchill's. If it is not pure Churchill, it has the marks of Churchill all over it.
(G, 26/5/83, X28, 4724-6)

[6.194] The terms were, the immediate withdrawal of German troops from Finland, the recognition of Soviet interests in the Balkans and the Straits through virtual control of Bulgaria, and the establishment of bases in the Bosphorus and the Dardanelles. If this is not Red Imperialism, I should like to know what it is.
(G, 11/4/83, X25, 4089-95)

d. The antecedent may express a desirable view or state of affairs. The connective if in this meaning is combined with only.

[6.195] ... how much better a place the world would be if only we were led by women rather than men.
(G, 8/6/83, X42, 6579-80)

e. The antecedent may express an alternative statement, usually stronger in wording and content than the one in the consequence.

[6.196] Many of the current economic ministers are believed to be due for "retirement" if not dismissal.
(STi, 24/4/83, X238, 37382-4)

f. The antecedent and consequent introduce two opposing or contrasting statements. Here conditionality is on the verge of adversativity.

[6.197] If Labour and Alliance agree on the need for working people to have a say in the running of their firms, they disagree violently on how this could be achieved.
(G, 3/6/83, X31, 5122-5)

[6.198] ... if we believe her, the Americans certainly do not.
(G, 7/6/83, X40, 6362-3)

g. The surface expression of conditionality may be used to embody

the statement of cause and effect.

[6.199] If one sees some signs of new realism in the OPEC air it is partly because its members have had to face the full horror of the market ...
(O, 23/1/83, X48, 7617-9)

In the Arabic corpus, most instances of rhetorical condition are of two types. The first corresponds to type (a) where the antecedent shapes the angle from which the statement in the consequence is to be viewed. In this example:

[6.200] wa 'iḏā kāna l-ī 'an 'abda'a min al-nihāyati, 'aw 'axtaṣira al-'asbāba... fa 'inna mawqifa al-'idārati al-'amīrkiyyati al-yawma, yakādu yakūnu al-mawqifa al-'isrā'īliyya nafsa-hu.
[And if I start from the conclusion or summarise the causes, (I shall say that) the standpoint of the American administration is identical with that of Israel.]
(Sh, 25/6/83, X147, 28266-70)

the conditional statement has a similar meaning to "To summarise, To conclude".

The second type of rhetorical condition is similar to, but broader and more diverse than, type (f) mentioned above. In this type, the antecedent and the consequent introduce two opposing statements, simultaneous actions or states, parallel views. The most frequent connective that expresses this function is "'iḏā", normally coupled with the defective verb "kāna". For example:

[6.201] 'iḏā kāna li-kulli šaxṣin ḥurriyyātu-hu wa huqūqu-hu al-fardiyyatu, fa 'inna li-al-fikri kaḏālika ḥurriyyāti-hi wa ḥuqūqa-hu..
[If everyone has his own personal freedom and rights, then intellectual thinking has also its own freedom and rights.]
(Hr, 23/5/83, X55, 10114-6)

[6.202] ..wa 'iḏā kānat al-'ārā'u taxtalifu fī 'asbābi hāḏihi al-ẓāhirati al-maraḍiyyati al-malḥūẓati, fa lā 'a<taqidu 'anna ṯammata xilāfan fī mu'addā-hā alladī yanū'u bi-hi laysa al-fardu fa ḥasbu bal al-mujtama<u..

> [And if opinions disagree as to the causes of this prominent and morbid phenomenon, I do not think there is any disagreement on its consequences, from which not only the individual but the society as a whole suffer ...]
>
> (Hr, 2/4/83, X47, 8589-93)

This flexible way of using rhetorical condition is common in the corpus.

6.12 Adversative

6.12.1 General Comments on Textual Functioning

The main function of adversative connectives is to signal that two knowledge configurations are, in the text world or in the general world knowledge, incompatible with each other. Thus their association together runs contrary to expectation. This relation, as with most connectives, can be exploited to relate bigger components, or 'chunks' of text, where the text producer introduces some arguments then pauses and starts to introduce counterarguments or incompatible, i.e. contrary-to-expectation, views, comments, outcomes or results.

Adversativity can be divided into two distinct functions: antitheticity and contrast. These are discussed below.

6.12.2 Antitheticity

6.12.2.1 Repertoire

The following connectives signal antithetic relations:

A. English

Simple: admittedly, alas, albeit, although, amazingly, and, anyhow, anyway, arguably, but, catastrophically, despite, even, fortunately, however, if, incredibly, ironically, irrespective, luckily, nevertheless,

nonetheless, notwithstanding, otherwise, paradoxically, rather, refreshingly, regardless, regrettably, sadly, still, strangely, suddenly, surprisingly, though, unfortunately, whatever, whichever, yet.

Compound: all the same, as well, at least, at any rate, but then, even if, even so, even though, oddly enough, in spite (of this), in any case, in any event, no matter, of course.

Correlate: whether ... or.

B. Arabic

Simple: lākinna, wa, lākin, 'innamā, bal, dūna, mahmā, ragma, fa, 'ayyan, 'iḏā,, lawlā, 'aw, bagtatan, siwā.

Compound:'illā 'anna, 'in, wa 'in, gayra 'anna, ma<a ḏālika, ma<a 'anna, wa law, ḥattā law, ḥattā wa law, <alā al-ragmi, li-al-'asafi, <alā 'anna, bi-dūni, ḥattā wa 'in, bi-al-ragmi, ma<a kulli, ma<a al-'asafi, <alā 'ayyati ḥālin, min dūni, ḥattā 'iḏā, kā'inan man, ma<a hāḏā, li-sū'i al-ḥaẓẓi, ma<a al-<ilmi 'anna, ḥattā 'in, 'illā 'an, bayda 'anna, bi-'asafin, bi-gayri, bi-ragmi, bi-kulli 'asafin, <alā al-<umūmi, min al-mu'sifi, min gayri.

Correlate: sawā'an...'am ['aw].

6.12.2.2 Textual Functioning

Antithetic connectives indicate that the subsequent proposition or sequence of propositions violate or run contrary to the normal

expectations about what the text world or the normal world looks like. This can be achieved in the following way:

1. The antecedent expresses a sufficient condition for the negation or non-occurrence of the content of the subsequent.

[6.203] Take productivity. Although the gains made during the labour shake-out in 1981 were certainly impressive, they merely offset the sharp productivity losses which were suffered as the economy moved into recession in 1979-80.
(DTel, 6/683, X116, 19385-9)

[6.204] ša<ar-tu 'anna-hā 'aṣbaḥat yatīmatan ragma 'anna-hā ta<išu wasaṭa ḥanāni al-wālidayni..
[I felt that she had become an orphan though she still enjoys her parents' presence and affection...] (Nb, 4/4/83, X1, 45-7)

2. Related to the previous function is one where connectives indicate exceptional, unexpected or undesired events, states or courses of action.

[6.205] Big new buildings may go up quickly, but their maintenance is poor.
(G, 11/4/83, X24, 3662-3)

[6.206] ..lā tanquṣu-nā al-qawānīnu wa lā al-rijālu, wa lākin tanqusu-nā al-himmatu wa al-'irādatu wa al-šu<uru al-ḥayyu bi al-wājibi.
[... we are not short of laws or men, but we are short of firm determination, will and a deep sense of duty].
(Hr, 23/6/83, X62, 11195-8)

6.12.2.3 Some Textual Patterns

1. A pattern that is often used for maximising the adversative relation in English works in this way. A proposition (or set of propositions) is presented and treated as an accepted truth by the use of such expressions as "it is true that". Then an adversative connective is used to introduce an exceptional, unexpected or

undesired comments, views, consequences, etc. For example:

[6.208] No-one should deny that the world recession has been an important element in explaining Britain's recession. But so have the following factors which go unmentioned in the Tory manifesto: a fiscal policy in Britain which the Organisation for Economic Co-operation and Development in Paris has calculated to be the most contractionary among the major developed nations; a monetary policy which targeted an irrelevant variable (sterling M3) in 1979-80, and forced British real interest rate up to lethal levels; and crucially an exchange rate policy which remained "cool" as Britain's trading competitiveness was devastated.
(DTel, 6/6/83, X116, 19345-58)

Note that the adversative nature of the relation in the previous text is displayed in the addition of unexpected "new factors" to the one mentioned first.

2. Some connectives conflate adversativity with a sense of conditionality. In English, connectives of this type are even if, even though, whatever, no matter (how), whichever. In Arabic, such connectives are "'ayyan" and "mahmā" [whatever, whichever] and those that are compounded with "ḥattā" (in the meaning of [even]): "ḥattā law", "ḥattā in", "ḥattā wa law", "ḥattā wa'in", "ḥattā 'iḏā". Observation of the structural patterns of these connectives in the Arabic corpus has suggested that normally each connective is followed immediately by a verb in the past form, though the meaning may be present or future. The connective "'ayyan" is followed by the defective verb "kāna" in the past. Examples:

[6.209] Whatever anyone thinks of the commission's individual proposals, no one can deny that the revision is long overdue.
(Ti, 13/11/82, X162, 26634-6)

[6.210] ..'ayyan kāna naw<u al-ḥurriyyati lā budda wa 'an yuṣāhibu-hā ba<ḍu al-ḍawābiṭi..
[... whatever the type of freedom is, it must be accompanied with some restrictions ...]
(Nb, 15/7/83, X21, 4134-6)

3. A pattern of adversativity that is related to the previous one involves the use of whether and its Arabic counterpart "sawā'an". In this pattern, a proposition (or set of proposition) is signalled adversatively against two (or more) alternatives. That is, the relation indicates that the content of a proposition obtains (or does not) in the text world regardless of the circumstances expressed in each of two other propositions. These are normally in sharp opposition to each other. In English second alternative may be the negation of the first, as in [6.211], but it need not be, as in [6.212].

[6.211] ... the Prime Minister is trapped into a June election whether next month's local election results go well or not.
(O, 24.4.83, X55, 8869-72)

[6.212] ... one factor common to almost all wars in history has been ... a cultural predisposition to war, whether this has been confined to ruling elites or widespread throughout society.
(G, 17/1/83, X10, 1401-5)

In Arabic, the two alternatives are combined by "'am" or, more recently, by "'aw" both having the meaning of [or]. Traditionally, the first alternative is introduced by the question particle "'a" but this seems to have receded. In the corpus only two occurrences of "sawā'an" (out of 16) have their first alternatives introduced by "'a". It should also be noted that the two clauses after "sawā'an" are used in the past though the verb time need not be.

[6.213] ..wa hiya mu<jizatun lā yastaṭī<u al-fikru al-siyāsiyyu al-yawma 'an yarsuma xuṭūṭa-hā sawā'an taxayyala al-ḥarba 'aw taxayyala al-'ittifāqa al-siyāsiyya..
[And it [i.e.having a unified Lebanon back again] is a miracle that no political mentality can draw up its lines whether it conjures up war or a political agreement].
(Sh, 25/6/83, X146, 28230-3)

6.12.3 Contrast

6.12.3.1 Repertoire

The following list comprises connectives that have been observed to have a contrastive function.

A. English

Simple: Conversely, instead, otherwise, paradoxically, rather (than), when, whereas, while.

Compound: by comparison, by contrast, for one thing, for another, in return, in another way, in other circumstances, in (sharp) contrast, in the face (of this, such), never mind, on one side, on the one hand, on the other, on the reverse [hopeful] side.

B. Arabic

Simple: wa, baynamā, 'illā, 'ammā, fīmā, faj'atan, 'iḏ, 'iḏā, badala, <adā.

Compound: fī ḥīni, badalan min, min nāḥiyatin, <alā al-<aksi, fī al-muqābili, min nāḥiyatin 'uxrā, min jihatin 'uxrā, bi al-muqābili, bi-<aksi, min jānibin 'āxara, muqābila ḏālika, xilāfan li-mā, bi-al-<aksi, <alā <aksi, fī muqābili, min jihatin, muqābila 'an, muqābila mā, <alā al-jānibi al-

'āxari, <alā hīni, fī muwajahati.

6.12.3.2 Textual Functioning

Connectives of this category signal a relation involving two (sets of) knowledge configurations that are normally in opposition in the text world. The opposition can take different forms:

1. "Genuine" opposition: where proposition A is the converse of, or placed in sharp contrast to, proposition B. This represents the main function of contrast and is signalled by the majority of connectives. For example:

[6.214] How is it when I attended a school reunion I could confront a face unseen over countless years and not only recall the name but attach with foolproof accuracy the correct initials, whereas my mind goes blank seeking to introduce a life time neighbour?
Conversely, I didn't seem to be quite so memorable ...
(G, 7/6/83, X39, 6213-21)

[6.215] wa kāna al-ba<ḍu waṯiqan wa mutafā'ilan, baynamā kāna al-ba<ḍu al-'āxaru <alā al-naqīḍi.
[Some were confident and optimistic, while others had the opposite attitude].
(Hr, 25/5/83, X58, 10862-3)

2. "Parallel" opposition: where propositions A and B are not in themselves inherently in opposition to each other; rather, they represent two parallel statements that are contrastive in the text world but not outside it. English connectives of this type are: on the one hand, on the other, in another way, in other circumstances, for one thing, in return. Arabic connectives: "min jihatin", "min jānibin", "min nāḥiyatin", "hāḍihi al-marrata", "fī al-muqābili". For example:

[6.216] The bill goes further by including any case where the policeman believes an arrest necessary to prevent loss or damage to property, an affront to public decency, or an obstruction to the highway. On the other hand, it does not extend arrestable offences ... to all offences bearing a sentence of imprisonment.
(G, 19/11/82, X4, 502-9)

[6.217] nataḥaddaṯu <an al-qawmiyyati al-<arabiyyati wa al-waṭani al-wāḥidi wa al-maṣīri al-muštaraki... min nāḥiyatin ’uxrā naḥnu nanqasimu dāxila al-waṭani al-waḥidi ’ilā ’aḥzābin wa tajammu<ātin wa ṭawā’ifa wa maḏāhiba.
[We talk about pan-Arab nationalism and one country and one common fate .... On the other hand we are divided in our own country into different parties, factions, groups and sects.]
(Sh, 20/6/83, X142, 27416-25)

The Arabic connectives of this type, it should be noted, signal contrast only in a broad sense. Often they conflate an additive sense of enumeration with the contrast.

3. Unexpectedness: This is an opposition based on the inclusion of an unexpected knowledge configuration that creates a sharp point of departure for the contrast. This function is more frequent in Arabic where it is signalled by such connectives as "’ammā" and "’iḏā (bi)", and by "bagtatan" used as a connective. The nearest English counterpart is the use of suddenly as a connective, though the few occurrences of this connective have been found more expressive of a variant viewpoint and hence considered orientative. The following excerpts exemplify the use of "iḏā (bi)" in Arabic.

[6.218] ..kun-tu ’ataṣawwaru ’anna-hu yu<iddu nafsa-hu li-ma<rakati-hi al-ḥaqīqiyyati <alā <aduwwi-hi al-ḥaqīqiyyi, fa ’iḏā bi-hi yatruku al <aduwwa wa yaqtulu nafsa-hu.
[I expected that he [the Palestinian fighter] was getting ready for his true battle against his true enemy, but all of a sudden he leaves his enemy in peace and kills himself].
(Ar, 27/6/83, X218, 37058-61)

4. Exception: This is a special type of opposition where a proposition is signalled as an exception in comparison to a norm in the text world. Connectives that signal this relation are except that (English), "'illā" and "<adā" (Arabic). Examples:

[6.219] Outside Eastern Europe, Mr. Andropov's direct foreign experience is limited except that as head of the KGB he clearly had an important watching brief on life outside the socialist world.
(G, 13/11/82, X2, 263-6)

[6.220] ..lam yakun laday-him mā yumkinu 'an yamnaḥ-ū-hu li-al-mujtama<i al-miṣriyyi 'illā mā qad yuḥaqqiqu la-hum 'agrāḍa-hum al-waḍī<ata..
[... there is nothing they could contribute to the Egyptian society except what would achieve for them their mean desires (and aims) ...]
(Hr, 27/1/83, X36, 6886-9)

5. Alternation: In this type, a view, event, state or course of action is opted for in contrast to another, more normal or expected, alternative. Such a function is signalled by the connectives: instead, rather, otherwise (English), "badala ('an), ḏālika", "badalan min ('an)" (Arabic). Note that these two Arabic connectives are equivalents of the English instead. No counterparts have been observed in the Arabic corpus for the other two English connectives.

[6.221] He [Nigel Lawson] had a duty to encourage new management, not frustrate it, to optimise and extend BNOC's achievements.... Instead, he has allowed a British success, Frank Kearton's BNOC, to deteriorate through malign neglect into the under-directed, under-financed and under-motivated Britoil that will creep out of the public sector this week with falling reserves and poor cash flow.
(G, 10/11/82, X1, 51-9)

[6.222] wa badalan min 'an yatawalla al-<arabu 'ittihama 'amirka bi-al-mas'uliyyati <an hadihi al-natijati al-mu'sifati qamat 'amirka bi-'ittihami al-<arabi wa tahmili-him al-mas'uliyyata fi mawti "al-mubadarati" wa mawti al-'ittifaqiyyati!
[And instead of the Arabs accusing America of its responsibility for these unfortunate consequences, America itself started to accuse the Arabs of killing "the initiative" and the death of the treaty].

(Nb, 15/7/83, X22, 4318-22)

6.13 Conclusion

This chapter has been concerned with outlining a textual grammar of connectives where emphasis is placed on their cohesive functioning and textual patterns. The term "connective" is problematic in linguistics. On the one hand, a number of labels have been employed to designate means of signalling logico-semantic relations in text; each reflects a certain aspect of connectivity and is commensurate with a particular linguistic persuasion. On the other hand, a particular label may refer to different entities, depending on how linguists envisage textuality. In both cases the result is a diverse and confused set of labels. For instance there is a confusion over differentiation between the categories of adverbs and conjunctions that is present even in authoritative codifications of language (such as dictionaries).

The term "connective" as used in this study is neutral in the sense that it does not require provocative reassessment of established categories, be they structural or functional. Generally, connectives refer to expressions that relate the subsequent text portions with the current one(s) and signal the semantic functioning of this relation. Structurally, these expressions may include a group of various syntactic constructions:

conjunctions, some adverbials (conjuncts, disjuncts and a few subjuncts), adverbial phrases with anaphoric reference and a number of clauses with conjunctive or disjunctive functions.

The logico-semantic relations that connectives signal are diverse. In their totality they reflect the organisational patterning of the text and throw the entire text into focus. An examination of the behaviour of connectives in the corpora has revealed a complex system of functionality that can be grouped into nine general categories: additive, comparative, alternative reformulatory, orientative, temporal, spatial, causal and adversatives. Each category may be regarded as a storehouse for means of creating various textual patterns that, collectively, contribute towards lending text its rhetorical organisation and subtlety of nuance.

We now need to examine the quantitative patterns of these categories within each corpus to establish numerically the textual impact of connectives and measure their functional variations. This numerical description, which has been referred to as the calculus of connectives (Chapter 1), will be the focus of the next chapter.

CHAPTER SEVEN

The Quantification of Connectives: A General Calculus of Observations

7.0 Perspective

One of the axioms of quantitative linguistic investigation is that, in textual measurement, we never prove, but only aim at stating the probability with which an event may occur (see, for instance, Herdan 1960, Dolezel 1969). This probabilistic view characterises our statistical profile of connectives. Since the observations are quantified within a text corpus, which is only a sample of the population, it follows that the observed characteristics represent random phenomena definable in terms of chance and probability. In other words, the numerical values that the observations generate, and that are displayed in sample frequency distributions, are liable to random fluctuations. These may, on the whole, range within a statistically admissible interval, and may only be chance variation of one and the same distribution. The existence of these variations, however, render the various statistical distributions, indices and values only approximate.

This statement, however, is by no means intended to compromise the accuracy of the calculus in reflecting the tendencies in the quantitative behaviour of connectives in the two corpora. For despite the fact that the occurrence of connectives in each corpus is of the nature of a probability, and thus a chance event, the regularity of the patterns is indicative of general linguistic norms that can be utilised in the contrastive description of connectivity across the two languages.

The description of the calculus has, for the sake of convenience, been divided into two parts: one is general in nature and the other is more specific in aim and direction. This chapter is mainly concerned with the first part. The starting point is a short review of some quantitative accounts of connectives available in the literature. The aim is to outline their scope and orientation and bring out their contribution and limitation. This review is followed by a description of the various statistical tools that are used in the various phases of the measurement, which collectively make up the calculus. Each tool is described from two dimensions: a) procedural, which considers the mathematical set-up of the measurement (its formula and expected result), and b) interpretational, which concerns the meaningfulness of the measure and the expected result to the linguistic analysis.

The computation of the first part of the calculus is made in two distinct phases:

1) Phase one is global; it aims at proffering a global account of the composition of the two corpora.

2) Phase two achieves a general profile of connectives. Here connectives are treated at a formal level, with little consideration of functionality. (For a detailed quantitative profile of functionality see Ch. 8).

The various observations that are calculated are first made of English and then of Arabic. This has the advantage of setting the two profiles that result from each phase of the computation in juxtaposition to each other and facilitate immediate comparison.

As explained in chapter 5 above, all phases of the computation are computer-aided. The use of the computer has assisted the

application of each set of measures in a controlled and systematic way to the text corpora. This, in turn, has ascribed three features to the statistical effort: accuracy, efficiency and speed, which, as a result, have achieved ample contribution to the validity of the calculus as a whole.

7.1 Some Quantitative Descriptions of Connectives

Studies of the statistical properties of connectives are limited in number and are restricted to investigating conjunctions: coordinators and subordinators. We would like to survey briefly four such studies, outlining their aims, procedures and results. These studies are conducted by Nebeska (1979), Smith and Frawley (1982), Kramsky (1983) and Whalley (1982).(1)

7.1.1 Conjunctions in Czech Newspaper Texts

Nebeska (1979) studies the quantitative characteristics of conjunctions in Czech newspaper texts and compares the results with those of journalistic as well as other styles as found in FDC (Frequency Dictionary of Czech).(2) Some of the results are:

a. The relative frequency of the conjunctions in the corpus does not differ from the data concerning the journalistic style of FDC. This implies that the relative frequency of conjunctions in one function style is not influenced by the extent (size) of a text.(3)

b. In terms of types, conjunctions in the corpus are more in number than those found in the scientific style and fiction styles of the FDC.

c. The tables of cumulative frequencies show that the 10 most frequent conjunctions cover nearly 87% of the total number of

conjunction tokens, and the 15 most frequent conjunctions cover 92%. These figures tally with FDC as a whole and the scientific texts. In texts of fiction the 10 most frequent conjunctions cover 90% of the total number of conjunction tokens.

d. Nearly 3/4 of the conjunction tokens are coordinators. This is explained by the high frequency of the coordinator "a". In terms of types, coordinators make up 59%.

e. The repertory and the order of the most frequent conjunctions in the corpus is constant in principle with the data in FDC and can, therefore, be regarded as a characteristic of the language.

f. Most of the findings concern the behaviour of coordinators and subordinators in terms of syntactic functioning. In general, the quantitative characteristics confirm that "for each syntactic function the most frequent conjunction is the fundamental one and that the other conjunctions are more or less synonymous" (ibid p.67). The fact is particularly applicable in the case of coordinators. In the case of subordinate conjunctions, the statistics conform to a certain extent with those of subordinate clauses. Nebeska then assigns a "fundamental" conjunction to each type of subordinate clauses, explaining their frequencies and positional features.

7.1.2 Properties of Conjunctions across Genres: Kramsky

Kramsky's (1983) stylostatistical investigation examines conjunction in three stylistic "strata": the style of fiction, the colloquial style and the (popular) scientific style. Each stratum is represented by a corpus of four texts, each comprising ca. 7,500 words.[4] Some of the statistical results are:

a. There are significant differences between the frequency of conjunctions in the colloquial style, which has the smallest number of tokens, as against the style of fiction and the scientific style. However, the difference in the relative occurrence between the latter two styles is rather small.

b. Much about the character of style is suggested by the relation between coordinate and subordinate conjunctions. The greatest relative difference in the frequency of coordinate and subordinate conjunctions is in the style of fiction, 72.59% (coordinate conj.): 27.41% (subordinate conj.). The least difference is in the scientific style (56.05%: 43.9%).

c. Differences in the relation of parataxis and hypotaxis are also detected in individual samples within each of the style of fiction and science.

d. As far as types are concerned, the scientific style has the largest number while the colloquial style has the smallest. This result correlates with the number of conjunction tokens (in a above).

e. The number of coordinate conjunction types is bigger in the scientific style and smaller in the colloquial style.

f. There are differences in the frequency of individual conjunctions in each style.

g. Within each style, some conjunctions appear to be statistically significant and characteristic of a certain author or of a certain scientific discipline.

7.1.3 Properties of Conjunction across Genres: Smith and Frawley

Smith and Frawley (1983) study conjunctions as cohesive ties,

using Halliday and Hasan's (1976) model of cohesion as a basis for their analysis of textual connectivity. The main aim of the study is to provide a systematic computational analysis of patterns of conjunctive cohesion in four genres of American English text: fiction, journalism, religion and science. The study is motivated by the argument that if conjunction is a crucial textual device that contributes to connectivity, and if connectivity is what makes a text a text, then it ought to be manifested differently in different genres of texts.

The corpus that is assembled for the purpose of the analysis consists of 16,000 word samples of fiction, journalism, religion and science, from the Brown English Corpus. Some of the results can be summarised as follows:

a. The distribution of coordinate and subordinate conjunctions reveal significant differences in the use of coordination over subordination for all genres (thus confirming one of Kramsky's results). However, the number of subordinate conjunction types is greater than that of coordinate conjunctions.

b. Religion and fiction have more conjunction tokens (1042 and 1143 respectively) than journalism and science (711 and 739 respectively), and are, therefore, more conjunctive.

c. Journalism and science use coordination much less frequently than fiction and religion. In fiction, coordination is twice as frequent as subordination, but in journalism they are almost equi-frequent.

d. The application of a X^2 test to all possible combinations of the fiction, journalism, religion and science samples shows that all genres differ significantly from each other at the .01 level, except

for fiction and journalism, which behave quite similarly. This latter result is explained on the basis that both fiction and journalism involve the creation of a sequential event line.

e. A comparison of the distribution of and and but in science and religion texts manifests some aspects of their rhetorical structure: it suggests a list or sequential argument structure in science, and an assertion and contrast structure for religion.

f. The distribution of individual subordinate conjunctions varies with genre types.

g. The distribution of conjunctions when they function cohesively (i.e. occurring extra- or cross-sententially and in sentence initial position) shows variations in relative frequency: science texts have the least relative frequency (3.2%) while, in comparison, fiction texts have the highest (10.4%).(5) Examined more closely, the distribution shows variation in the frequency of cohesive coordinating and cohesive subordinating conjunction in each genre.(6)

h. Distribution of semantic categories (two in the case of cohesive coordinators: additive and adversatives, and four in the case of cohesive subordinators: additive, temporal, adversative, causal) varies considerably in each genre.

7.1.4 Density of "Rhetorical" Connectives

The fourth study of the statistical properties of connectives is conducted by Whalley (1981). The context for this study is an attempt to construct an advice system for use in the production of low resource educational text. Such a system, Whalley suggests, can assist in proffering a comparative analysis of text complexity, or

some measure of the development or emphasis given to a particular topic.

Whalley's analysis is concerned with one aspect of cohesion in text: the use of "rhetorical connectives" or "items of conjunction". The aim of the study is to outline some portion of the text writer's intended rhetorical structure as it is likely to be perceived by the reader, who relies on the explicit signalling clues available. One such clue is the large store of connective operators that can be used to indicate precisely where each new idea fits into the discourse structure.

The main thrust of Whalley's study involves a statistical analysis of the density of rhetorical connectives between different parts of a text or texts. A limited list of coordinators and subordinators is prepared and their statistical density count is established across a range of texts. The corpus that is analysed comprises 33 samples, each of about 1,000 words, taken from three levels of an Open University course. The results of the count indicate that the coordinator subordinator distinction is capable of providing useful comparative data. Whalley suggests that the subordinators "may provide an index of syntactic complexity" and that (intersentential) coordinators" can indicate the general form of discourse within a text" p.56).

Further, the study investigates the relation between rhetorical connectives and paragraph structure. This is based on a statistical analysis of the paragraph distribution of connectives for several texts. This part of the study is however not conclusive; the figures in the statistical analysis is left without discussion.

The analysis of density of connectives is extended further to a study of the development of important topics in texts. This is "effectively a selective word frequency analysis of a concordance output, and represents another attempt to see whether simple statistical analysis can provide useful information" (p.57). This analysis is concerned with two complete Foundation Level science units (each of 80-90 thousand words) and reveals certain differences in the frequency of coordinators; no difference in the pattern of subordinators is detected. Again this part of the study is not well-discussed and the emphasis is placed on the statistical graphs. Whalley admits that it is not clear to what extent his 'surface' analysis can provide an entirely sufficient measure of text complexity, and that his measures are too simple to provide more than a source of comparative data.

7.1.5 Concluding Remarks

These statistical analyses, valuable for the type of task each sets out to study, suffer from some obvious shortcomings, some being admitted by the researchers themselves. A number of such weaknesses are here noted; some are qualitative while others are quantitative in nature.

1. One lamentable deficiency in these studies is their lack of specificity regarding the nature and function of connectives. All efforts have been concentrated on a limited list of coordinators and subordinators, leaving out other types of connectives (probably because these other types have not been regarded as connectives in the first place). The result is a too restricted conception of textual connectivity which, in general, has rendered the description

too narrow if not blurred and inexact.

2. With the exception of Smith and Frawley's account, these studies have neglected consideration of the various functions of connectives. The statements made are thus too generalised.

3. Also left out from the discussion and calculation are all types of multi-word connectives. Whalley makes a specific admission of this weakness while the other studies fail to note their existence.

4. Generally, the statistical procedures used in the analyses are limited. Comparison has often been based on absolute figures and there is a lack of such quantitative features as the analysis of concentration, repeat rate or growth of connectives in each corpus.

These limitations have the effect of rendering the quantitative descriptions only approximate. However, it should be stated that while the operational simplifications in these studies weaken the absolute rigour of the analyses and conclusions, they stop far short of vitiating it entirely. It suffices to accept them as methodological deficiencies that have to be avoided in the present work.

7.2 Measures used in the Calculus of Connectives

7.2.1 Preliminaries

This section is intended as an introduction to the subsequent statistical analysis (here termed the calculus of connectives). The aim is to expose ab initio the types of measurements that constitute the calculus and discuss their main characteristics prior to the discussion of the calculations in each language. These measurements are corpus-based and represent what Herdan (1962, p.18) calls

"running text statistics". There are admittedly far more types of language statistics that can be employed here; however, in trying to be specific in our account of the statistical properties of connectives, we have to use those techniques that collectively enable us to arrange the vast numerical material in an orderly fashion and to present it as an organic whole. The techniques used here are: frequency distribution and relevant measures, repeat rate, growth rate and measurement of intervals. These are defined in the next few sections.

7.2.2 Frequency Distribution and Related Measurements

The first measurement we apply is calculation of the frequency distribution of connectives. This involves counting the number of occurrences of each connective in the whole length of text corpus. Two types of frequency lists are then produced: in the first key words are arranged alphabetically and in the second according to descending frequency. The relevant calculations give the absolute, relative and cumulative frequencies for each connective type.

Closely linked to these lists is the rank distribution. This refers to the ordering of the frequencies whereby connectives with the same number of occurrences are grouped in one class. The rank distributions are useful as a basis for other calculations, such as the entropy.

These calculations are useful in deducing probabilities of occurrence of connectives in the corpus (which is, as discussed in Ch. 5, only a sample drawn from a population). In general, a phenomenon's sample relative frequency will differ from the same phenomenon's population relative frequency or probability (see the

discussion in Knowles (1981 pp.167ff). However statistical techniques are used to reduce these differences and to ensure that lack of sufficient knowledge of the population does not hinder the deduction of inferences on the basis of accurate calculations of its sample. We accept that it is impossible to compute from sample relative frequencies a single 'point' value for the corresponding population probability. It is therefore practical to set an interval of values within which the population parameter is to be encountered a substantial percentage of the time (i.e. 90%, 95% or even 99%). Hence our use of confidence level in some computations.

7.2.3 Measures based on type-token mathematics

A number of measures based on type-token mathematics are applied here to the computation of connectives. Some are suggested by Herdan (1957, 160, 1967), but most of them are proposed in Mistrik (1973) and elaborated and applied in Knowles (1981). Collectively, these measures, as applied here, offer a statistical profile of the textual behaviour of connectives in the two corpora.

1. The type-token ratio:

This is a standard measure in statistical linguistics. It simply refers to the ratio of the vocabulary in a given text, or sample from it, to the total number of words comprised in the text of the sample. For texts of equal length the type-token ratio (TTR) provides an adequate measure for a comparison of their vocabulary richness.

However, this quantity changes, in general, with the size of the text. Vocabulary increases with text length, but by no means proportional to it, in such a way that the quantity decreases, on

the whole, with increasing sample size.

This fact has led some scholars (for instance, Herdan 1957, Carroll 1967) to argue that the type-token ratio cannot serve as a characteristic of vocabulary richness, which must be independent from the text length. Herdan proposes that the logarithmic type-token ratio, i.e. log type-log token, remains sensibly constant for samples of different size from a given text and, hence, is suitable to serve as a style characteristic. Carroll believes that there is considerable support for supposing that the theoretical population is distributed lognormally.(7) Other scholars use the log-log type/token and consider it a more adequate measure for vocabulary richness (see applications in Tuldava 1977 and Knowles 1981).

These measures will be used in the calculus, first as indicators of connective type extensity in the two corpora, and later as part of the calculation of other quantitative characteristics of connectives (see below).

2. Concentration

The concentration of connectives is expressed as the number of non-hapaxes divided by the total number of connective tokens. The index is directly influenced by the number of connective hapax legomena in the corpus. The bigger is the number of hapaxes, the smaller is the number of repeated connectives and, therefore, the lower is the concentration index. Conversely, the smaller is the number of hapaxes, the bigger is the size of repeatedness, thus yielding a higher concentration. Accordingly, the size of concentration is closely related to the size of repeatedness, particularly to the size of the share that each non-hapax has in the

token mass.

3. Exclusivity

This quantity represents the size of the share in token mass reserved by the hapaxes. It is obtained by the simple formula

$$E = \frac{A}{N}$$

where hapaxes are divided by the total number of tokens. The index is closely related to the concentration index since the sum of both indices represents the TTR of connectives, i.e. the index of type extensity within the corpus.

4. Hapax probability

The probability that a "draw" of one connective from an "urn" containing all connective types will yield a hapax legomenon is expressed by the formula

$$Pa = \frac{A}{V}$$

Obviously, this formula is sensitive to the number of hapaxes. The bigger the number, the higher is the probability. The complement of this index gives the probability of the non-hapaxes, which coincides with the consolidation factor (see below).

5. Variegation

The computing of the hapax probability of connectives helps the calculation of the index of variegation, which is simply

$$G = Pa * 100.0$$

This index indicates the size of diversification in connective types. The index ranges from 0 (the case of minimal variegation when hapaxes are equal to naught, A = 0) to 100 (the case of maximal

variegation when hapaxes are equal to the number of connective types, A = V).

6. Consolidation factor

This factor is calculated by using this formula

$$Cof = 1 - \frac{A}{V}$$

where A refers to the sum of all hapaxes and V to the number of types. This factor coincides with the probability measure of the non-hapax connectives and is, therefore, sensitive to the size of hapax legomena within connective types. It ranges from 0 (the lowest consolidation factor indicating that all types are hapaxes) to 1 (the highest consolidation factor indicating the absence of hapaxes). The index is thus indicative of intensity of use of types.

7. Density

The density of connective use is obtained by the formula:

$$D = \frac{ld(N)}{1 - \frac{A}{N}}$$

where ld(N) is the binary logarithm of connective tokens. This quantity, like concentration, is influenced by the number of connective hapax legomena. If the hapaxes are large in number, density of connectives is small and the converse is true. This is because the more hapaxes there are, the smaller is the share that types in general and non-hapaxes in particular can get from the token mass, and therefore the smaller is the density. On the other hand, if the number of hapaxes is small, non-hapaxes will achieve a

larger token coverage.

8. Gravity

This measure indicates a relation between the number of connective tokens and the number of connective hapax legomena. The index is obtained simply by dividing the total number of tokens by the number of hapaxes. The quantity represents the hapax occurrence rate within connective token mass. One direct application of this index is for comparing the extent of connective saturation within the two corpora. A high index represents a bigger number of hapax legomena and therefore slower saturation. Conversely, a low index indicates a smaller number of hapax legomena and therefore a faster saturation.

9. Rhythmicality

This index represents the extent of repetitiveness of the various non-hapax connectives. This is clear in the way the index is obtained:

$$R = \frac{N - V}{V - A}$$

The function N - V represents the size of repeatedness and V - A denotes non-hapaxes. The index therefore shows the share of repetitiveness that each non-hapax has within the corpus. This measure is useful for the purposes of comparison. A higher index represents a high share of repeatedness due either to a small number of types and/or non-hapaxes. A lower index, on the other hand, indicates a small share of repeatedness caused by a large number of types.

10. Type occurrence rate

The type occurrence rate of connectives is interpretable as the lexical distribution mean of the connectives. It is represented by the token-type ratio, or, alternatively, by the formula

$$T = 1 / \frac{V}{N}$$

It indicates how many connective tokens we may have before we encounter a new type.

11. Stereotypicality

This index measures how stereotypical connective types are in a corpus. It is obtained by the formula:

$$S = (N - A) / (V - A)$$

The quantity strips the connective tokens from all hapaxes and divides them by the non-hapaxes. Accordingly, if the size of N is big and the size of V is small, the index will be high and the set of connectives can be characterised as highly stereotypical. The converse is true. This measure is similar to the token-type ratio without the hapaxes; one should, however, make adjustment for the fact that hapaxes will always occur. This explains why we shall use both measures.

12. Predictability

The predictability index of connectives is the complement of the type-token ratio and is calculated simply by:

$$P = 1 - V / N$$

Since the type-token ratio indicates extensity and richness, its complement indicates the size of predictability of connective types. It follows that the extent of P (predictability) is on converse

terms with the size of V (connective types) provided that N is the same. Accordingly, an increase in the size of V will reduce the predictability of its members, the minimal predictability index being 0. On the other hand, a decrease in V size will increase the index (maximal predictability being 1).

13. Lexical efficiency and lexical usage

These two measures are related to the entropy of connectives (see below). Lexical efficiency is calculated by dividing the entropy of connectives by the binary logarithm of connective types, while lexical usage is computed by dividing the entropy by the binary logarithm of connective tokens. Both indices are useful for comparing some further quantitative properties of connectives in the two languages.

7.2.4 Entropy and Redundancy

Entropy is a measure of information. It is expressed in units termed bits (contraction of binary digits) and represents the amount of uncertainty or statistical randomness attached to a set of objects, here connectives. According to Herdan 1956, 1966, (see also Shannon 1948, 1951, Shannon and Weaver 1949) there are two types of entropy that can be calculated within the linguistic code. First, H' is a measure of information about a set of symbols without regard to their statistical distribution, that is without regard to their having definite a priori probabilities, or, in other words, under the condition of their equidistribution. It, therefore, represents a measure of the uncertainty in the choice of members of a particular set of linguistic units if all these members are

equally likely, that is if there is no constraint whatsoever in choosing a member.

The other measure of entropy, H, takes into account the actual probability distribution of the linguistic symbols (here connectives). This is interpretable in the following way. Since certain members of a set of symbols (connectives) occur consistently more often than others, which leads to a gradient of probabilities characteristic of the language (English or Arabic), then the use of the symbols, apart from chance fluctuations, are governed by that gradient of probabilities, and should be reflected in the computation of the entropy.(8)

The quotient of these two entropies, i.e. $h = H/H'$, is called relative entropy. It represents the influence of weighting ranks of the symbols by the actual probabilities, or roughly, the influence of the statistical distribution of the symbols. Stated differently, the relative entropy indicates to what extent the uncertainty in the use of a symbol has been reduced through the gradient acting as a constraint upon the free or chance use of the symbols. The complement of the relative entropy, i.e.

$$R = 1 - h$$

is a measure of the redundancy of the code.(9) This measure refers to the property which

> " ... enables us to use the stability of the relative frequencies for making guesses as to missing parts of the message with a reasonable degree of expectation to be correct." (Herdan 1956 p.165)

The measure of redundancy is therefore closely associated with the entropy: it is greater the smaller H, and thus, the greater the

state of order or the degree of structure in the language. Conversely, the smaller the calculation of redundancy, the greater the size of H, and hence, it is a sign of lack of structure (and thus of disorder and chaos).

7.2.5 Measure of growth and Tuldava's index

The measure of growth reflects the dependence of vocabulary size on text length. The determination of this measure as regards the types and tokens of connectives is necessary if we would like to make a differential statement regarding the degree of saturation of types of connectives in English and Arabic. To achieve the calculation of growth we shall make us of Tuldava's proposals made in his 1977 paper. We believe that Tuldava's formula is simple, practical and can describe the growth of vocabulary with sufficient exactitude.(10)

In an earlier paper (1974), Tuldava stipulated the fraction-linear function of types (L referring to types and N to tokens).

$$L = \frac{a\ N}{N + b}$$

mainly for the purpose of resolving some stylostatistical problems. "In this formula a is the asymptote which points to the limit of L if N increases infinitely" (1977 p.29). This constant may be interpreted, and is indeed declared in the statistical tables of growth, as the index of lexical richness of a given text. In the case of this project it refers to the richness of connective types in either corpus. The constant b (or, more precisely, the relation a/b) determines the rate of growth of vocabulary L in respect to the growth of the length of the text N. In this project, this describes

the growth rate of connective types in relation to the increase in connective tokens. Tuldava believes that "both of these constants may be considered to be stylostatistical characteristics which may serve as differentiating indices for the comparison of the lexical structure of various texts" (loc. cit.).

The formula mentioned above presupposes a linear interdependence between the average frequency of occurrence of words (N/L) and the size of text N. Such a relation can be observed within certain intervals of text (in many languages, Tuldava states, between N = 3,000 and N = 30,000). "In this interval the given formula may be used in order to determine the growth of vocabulary and the potential richness of the vocabulary of the compared texts" (loc. cit). However, in general the relation is still non-linear, and for the purposes of determining some regularities of the generation of texts, Tuldava proposes the use of the relationship between the double logarithms lg lg N/L and lg lg N, which, he maintains, remains practically linear on any text size up to $N = 10^6$. As a result of his calculation, the linear function

$$\lg \lg N/L = A + B \lg \lg N$$

is modified into a more appropriate form by eliminating the logarithms on the left hand side, thus rendering this formula:

$$N/L = N^{a(\lg N)^b}$$

or

$$L = N^{1-a(\lg N)^b}$$

where $a = 10^A$ (antilog A) and $b = B - 1$

This function can be applied for extrapolation. The predictions that it can render have been tested and found that they coincide

almost exactly with the value observed empirically with respect to the size of vocabulary. An extrapolation towards decreasing text sizes also yields acceptable results.

However, extrapolating beyond $N = 10^6$, the calculation results in a limit, i.e. the function has a maximum.(11) For instance, Tuldava found that in the case of the frequency dictionary of English $N_{max} = 10^{11}$ and the maximum value of the vocabulary size is consequently L = 800,000. Nevertheless, Tuldava admits that the coincidence of the maximum of the function and the limit of the vocabulary size cannot be hypothesised.

Since the size of connective tokens do not go beyond the maximum value that the function suggests, we can safely state that the function is able to describe the dynamics of connective growth in our two corpora with a high degree of efficiency. The descriptive power can help us to extrapolate to huge sizes of connective tokens. We would like to make two points in this connection:

a. Tuldava's function has been used to compare global text tokens and types. To my best knowledge, this study is the first to analyse the growth and make extrapolation of the size of a particular linguistic unit, the connective, in a corpus of text, using Tuldava's formula.

b. This study, again, is the first to apply Tuldava's formula to, and so receives further support from, a corpus of Arabic text. Other languages that it has been used for include English, Polish, French, Czech, German, Romanian, Estonian, Latvian and Kazakh.

7.2.6 Repeat rate

The repeat rate of a connective is simply the square of that

connective's probability (Herdan 1960, Knowles 1981). This measure represents the probability that two connectives chosen at random from a great mass comprising the connectives of the entire corpus will yield the same connective. The mathematical expression for this is straightforward. If the probability of a particular connective is Pi, the repeat rate is Pi^2.

The repeat rate of a connective can be extended for all connectives. It now represents the probability that two random "draws" from the "urn" containing the vocabulary of the entire corpus will be connectives. In this case we sum over all the repeat rates for the different connectives, say n in number:

$$\text{Repeat rate} = \sum_{1}^{n} pi^2$$

Another application is to find out the repeat rate of a connective denoting a particular semantic category. This represents the probability that two connectives chosen from a mass of connectives turn out to belong to the same functional category, i.e. both additives, adversatives, etc.

7.2.7 Measures of Interval

Two types of interval measurements are used. Both aim to measure the distance between each two successive occurrences of connectives. For lack of differentiating terms and in order to make a proper distinction between the two measures, we shall refer to the first one as the measure of gap and the second as the measure of distance. The measure of gap is based on a similar measure used in the Russian word count (Josselson 1954) while the other is the application of the formulae proposed by the Russian scholar Levin in his 1967

paper.

In the analysis of distance, the corpus is regarded as a continuous sequence of sites or places occupied by words, i.e. the units in terms of which the gap length is expressed. Hence the length of a distance between two occurrences of a connective is measured by the number of such sites in the linear sequence between the occurrences in question, exclusive of both occurrences. Thus for two immediately successive occurrences of connectives, we say that the interval length is zero; if they are separated by one site, the interval length is 1, and so on. In the case of simple connectives the gap length is measured from the site where the connective is positioned to the next. In the case of multi-word connectives, it is measured from the site occupied by the connective core; any component of that connective (the peripheral component, i.e. the rest of the words that make up the multi-word connective, see Ch. 5) is counted as part of the gap.

Below is a discussion of each of the two types of measurements of interval.

1. Measure of Gap

The aim of this measure is to provide a numerical distribution of distance in terms of intervening words between repetitions of connective tokens. The distribution enables us to draw comparisons between gap lengths in the English and Arabic corpora, specifying the number of ranks, the shortest and longest observed distances and the average gap length.

This empirical distribution of gap length assists in describing patterns of repetitiveness of connectives in each corpus. The variations that result from contrasting the patterns can be assessed

against other findings in order to determine general trends in text organisation within each language. In particular, we shall consider the question whether the distribution of connective repetitiveness is in conformity with language-specific textual patterns of cohesion.

2. Levin's Measure of Interval

Levin's measure of the ordering of the distribution of a determined symbol in text is expounded in his 1967 paper.[12] The symbol may be a phoneme, a syntactic construction, an intonation curve, a rhythmical form (in poetry), etc. The measure aims to compute the manner of repetitiveness of the symbol. More specifically, it shows how compactly, or conversely, how diffusely this element is distributed in text, i.e. whether its occurrences in text have a tendency to concentrate at separate places divided one from the other by great distances, or conversely, is there a tendency towards an equal diffusion throughout text. Hence Levin's measure, unlike the measure of gap discussed above, computes the characteristics of intervals independently of the frequency of the element or symbol under investigation.

To elaborate Levin's arguments further, let us suppose there is a certain finite sequence consisting of symbols of two sorts a and B. Suppose also that a is a symbol that interests us, and B denotes all other symbols. The question under study, then, is: how compactly is the symbol a distributed in this sequence?

Let the sequence consist of n occurrences of the symbol a and m of the symbol B. If we number all occurrences of a from left to right we will obtain a notation of the sort:

$$B\ B\ a_1\ \ B\ B\ B\ B\ a_2\ \ B\ a_3\ a_4\ B\ B\ \ldots\ a_n\ B$$

Let us denote by d_i ($i = 1,2, \ldots, n - 1$) the distance between a_i and a_{i+1}, i.e. the number of symbols B which separates them, and $0 < d_i < m$. Let us introduce, in addition, the value d_n, setting it equal to the total number of B, preceding a_1, and following after a_n. Then

$$\sum_{i=1}^{n} d_i = m.$$

Levin considers as maximally compact (with regard to $\underline{a}$) a sequence for which all $d_i = 0$ apart from one equal to m. For instance, a sequence of the sort

B B a a a a B B B

is maximally compact. A maximally diffuse sequence would be one in which $\underline{a}$ is distributed equally, that is with

$$d_1 = d_2 = d_3 = \ldots d_n = m / n$$

(assuming that m / n is an integer; otherwise an equal distribution is completely impossible).

Levin goes into minute details in working out his formulae. He then illustrates the application of these measures in concrete examples, calculating compactness of particular words, sentence lengths and metric measures. It is interesting to note that what Levin argues for is the type of repetitiveness of intervals of a particular symbol. What he does not tell us, however, is what would happen if one wanted to compute the repetitiveness of members of a category of symbols (e.g. a class of words, connectives) rather than instances of one single symbol (word). That is what our calculus will try to find out.

7.3 Global Statistical Profile of the Corpora

7.3.1 Statistical Profile of the English Corpus

7.3.1.1 Frequency and Rank Distributions

According to the statistical account obtained from the OCP runs, the English corpus comprises 256,560 tokens. The number of types is 20,064 of which 9,548 are hapaxes.

A close examination of the rank distribution shows that out of 294 ranks, the first ten, occupied by one type each, representing the highest frequency types, achieve more than 25% of total token coverage. In terms of types, however, these ten ranks achieve only 0.05% of the size of vocabulary (i.e. of the total number of types). These figures correspond neatly to the frequency figures displayed by the statistics of three corpora: the Brown (see Kucera and Francis 1967), the LOB (Johansson and Hofland 1982) - each of these comprises a million word tokens - and the English journalistic corpus (Alekseev and Turygina 1974; this corpus comprises 200,000 word tokens). In the Brown corpus, the types occupying the first ten ranks achieve 24.256% token coverage; in the LOB it achieves 24.521% and in the English journalistic corpus it achieves 26.44%. The last figure is higher because the types in this frequency list represent the lemmatised forms of words and this fact affects two ranks within the top ten: combining in one the frequencies of "a" and "an", and in the other the frequencies of all forms of the verb "to be".

The next top 10 ranks in our corpus achieve collectively a token coverage of only 6.28%, bringing the token coverage of the 20 most frequent types to 31.43%. Again these figures tally with those in

the three frequency lists. In the Brown, the second ten ranks constitute 6.779% of total word tokens, bringing the total coverage of the top 20 to 31.035%. In the LOB, the second top 10 represents 7.162% of total word tokens, a slightly higher figure which brings the total token coverage of the first 20 ranks to 31.683%. In the English journalistic texts, the second top 10 ranks constitute 6.20% of token coverage and the combination of the top 20 ranks is, therefore, 32.647%.

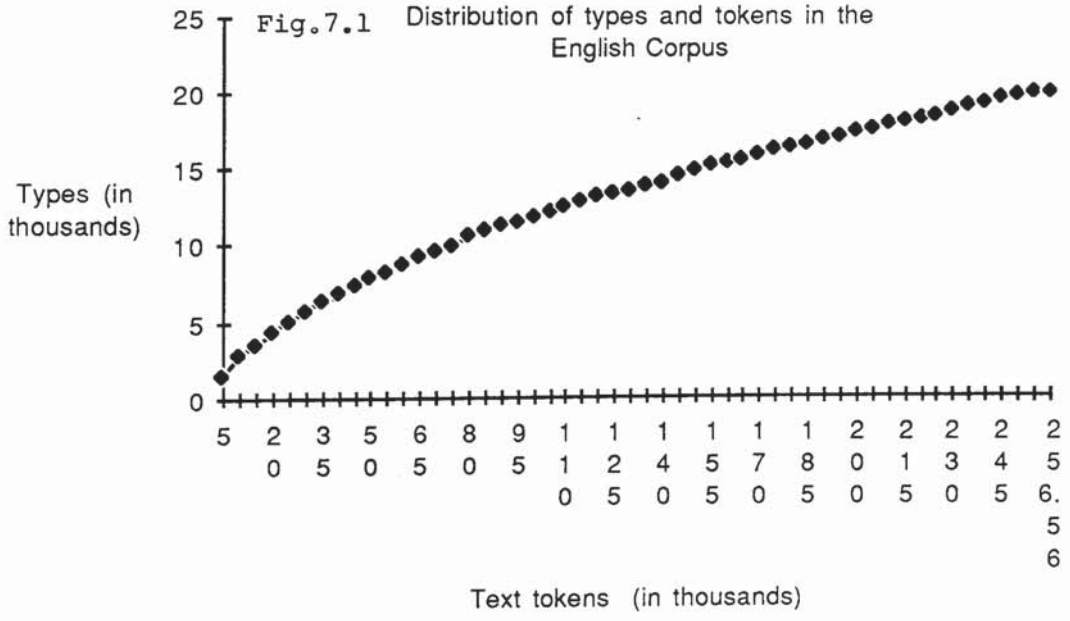
7.3.1.2 Type-Token Measurements

The type-token ratio in our corpus is 0.078 noticeably higher than the same figure in the Brown and LOB, both approximately 0.05. The reason is obvious and has been mentioned in 7.2.3 above. The type/token ratio is sensitive to text size and tends to decrease as text size increases. (We shall not make comparisons with the frequency lists of the English journalistic texts; the fact that they are based on lemmatised forms makes comparability erroneous.) However, the logarithmic type-token ratio is nearly similar: 0.796 in our corpus and 0.783 in the Brown. The double logarithmic type-token ratio gives the same figure in both corpora: 0.864 (reduced to three decimal places). Hence it seems that despite the generality of the Brown and LOB corpora (in the sense that each represents 15 genres of prose) and the specificity of our corpus (being based on newspaper texts), the extensity and richness of vocabulary, if text size is neglected (which is the purpose behind using the logarithmic and double logarithmic ratios), is similar in all three corpora. The relation between the types and tokens in our corpus is shown in Figure (7.1).

A number of other measurements have been calculated using the size of the vocabulary and the length of text. These are summarised

Concentration:	0.040969
Consolidation factor:	0.523091
Density:	18.6662
Exclusivity:	0.0373519
Gravity:	26.7724
Hapax probability:	0.476909
Non-hapax probability:	0.523091
Lexical efficiency:	0.707938
Lexical usage:	0.563172
Type-token ratio:	0.0783209
Predictability:	0.921679
Rhythmicality:	22.497
Stereotypicality:	23.497
Type occurrence rate:	12.768
Variegation:	47.6909

Table 7.1 Indices based on type and token measurements of the English corpus



in Table (7.1). The concentration of the vocabulary within the corpus is 0.041. This figure is noticeably low compared to the exclusivity index, which is computed at 0.037. The figures are justified on the basis of the rather high number of hapax legomena. Indeed the probability of the hapaxes is 0.477, only slightly lower than the probability of the non-hapaxes: 0.523. The index of variegation, indicating the size of diversification within the corpus vocabulary, is 47.7%. The gravity index is calculated at 26.77.

The type occurrence rate is approximately 13. This refers to the number of tokens we have before we encounter an new type. The index of stereotypicality of vocabulary is 23.5. The index of rhythmicality of vocabulary repeatedness is 22.5. The intensity of the repeatedness of types is calculated at 0.523, which is the consolidation factor. Lexical efficiency index is 0.70793 and index of lexical usage is 0.56317.

7.3.1.3 Sentence Length

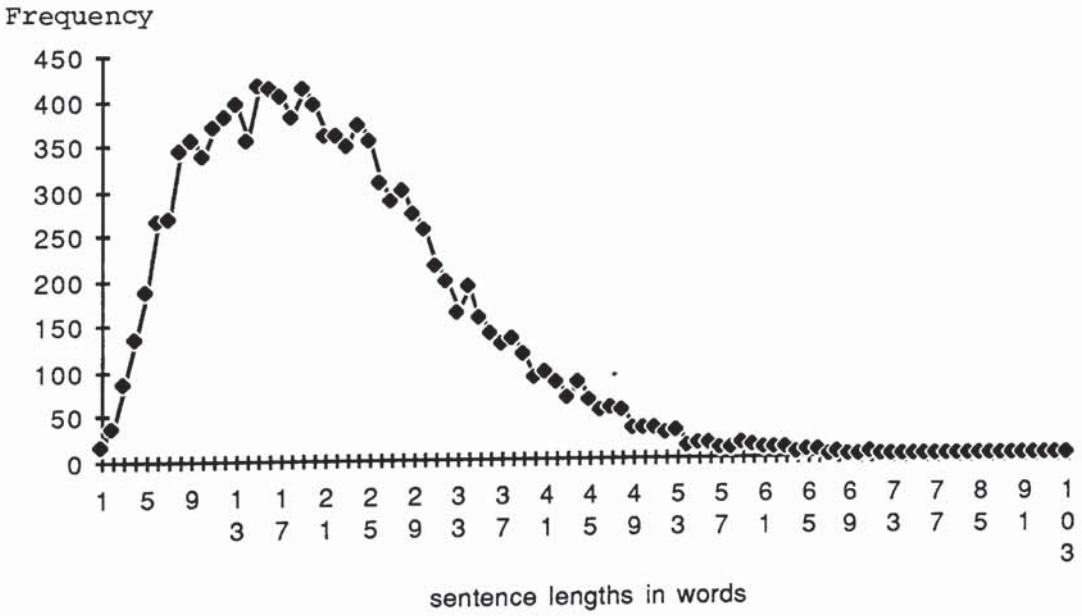
Sentence length has been calculated and the results are in Table (7.2) (full analysis of sentence length is in Appendix 10). There are 11,671 sentences in the English corpus with an average sentence length of about 22 words. The use of a confidence index shows that in 95% of cases the sentence length will range between 21.76 and 22.20. The standard deviation is 12.17.

The longest sentence in the corpus comprises 103 words, which occurs only once. The shortest sentence is made up of one word only and has a frequency of 16 within the corpus. The most frequent sentence length is 15 words, having 416 occurrences and a relative

Total number of words	256560
Total number of sentences	11671
Average sentence length	21.9827
Variance	148.0101
Standard deviation	12.1697
Coefficient of variation	55.3603%
Standard error	0.1126
Lower 95% confidence limit	21.0762
Upper 95% confidence limit	22.2035

Table 7.2 Sentence length in the English corpus

Fig. 7.2 . Distribution of Sentence Lengths in the English Corpus



frequency of 3.56%. The coefficient of variation is calculated at 55.36%. Figure (7.2) shows the frequency distribution of sentence lengths in the corpus.

7.3.1.4 Paragraph length

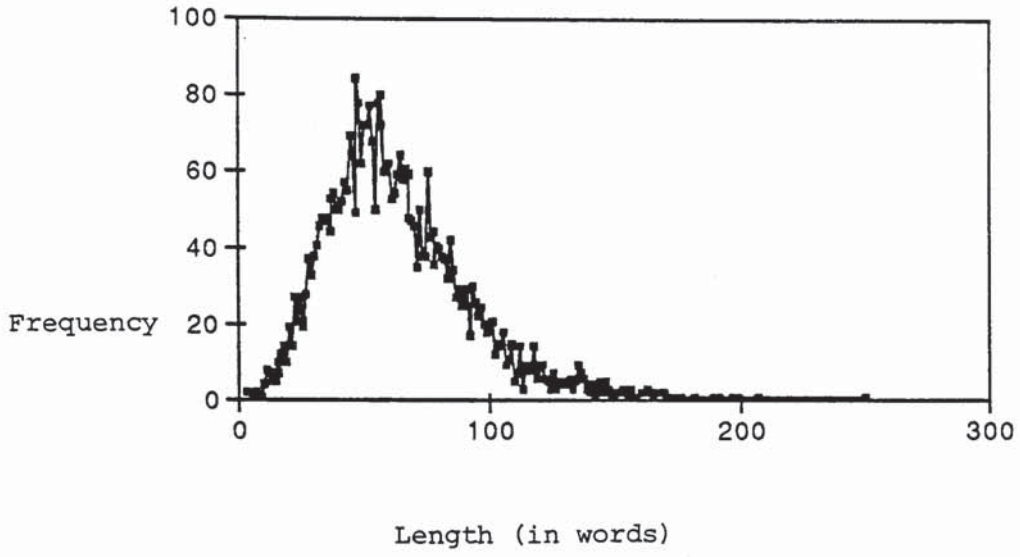
The measurement of the number and size of the paragraphs in the corpus has given the results in Table (7.3). A fuller description is given in Appendix (12). The corpus has a total number of 4083 paragraphs with an average paragraph length of approximately 63 words and a standard deviation of 28. The smallest paragraph comprises 3 words and has two occurrences. The longest paragraph

Total number of paragraphs:	4083
Total number of words:	256559
The smallest paragraph had	3 words
The largest paragraph had	250 words
The average paragraph length is	62.84 words
Variance is	795.00367
Standard deviation is	28.02023
Coefficient of variation is	44.88%

Table 7.3 Paragraph length in English

runs into 250 words and occurs only once. The most frequent paragraph size is 47 words which occurs 84 times and has a relative frequency of 2%. The coefficient of variation in paragraph sizes is 44.88%. Figure (7.3) gives a visual description of the distribution of paragraph lengths.

Fig. 7.3 Distribution of Paragraph Length in the English Corpus



7.3.1.5 Calculation of Growth

One other measurement we applied is the calculation of growth and extrapolation. Intervals at which growth is examined is set equal to 5,000 tokens, and types are computed at each interval. Then we applied Tuldava's formula to extract the richness and growth rate indices which give the bases for extrapolation. The results (see Table 7.4 and Figure 7.4) show that Tuldava's richness is 0.008872 and the computed growth rate is 0.007135. The figures of the expected number of types within each interval tally to a considerable extent with the observed number of types at most points. This gives support to the rationale and theorisation behind Tuldava's formulae and to their applicability to the measurement of the relation between vocabulary size and text length.

Tuldava's richness index: 0.008872
Connective growth rate: 0.007135

Actual Tokens	Actual Types	Expected Types
5000	1633	1690
10000	2736	2746
15000	3610	3613
20000	4393	4370
25000	5117	5054
30000	5788	5681
35000	6379	6265
40000	6883	6813
45000	7426	7331
50000	7894	7824
55000	8357	8296
60000	8871	8747
65000	9345	9182
70000	9771	9601
75000	10118	10006
80000	10573	10398
85000	10955	10779
90000	11261	11149
95000	11569	11509
100000	11843	11860
105000	12215	12203
110000	12592	12537
115000	12942	12864
120000	13224	13184
125000	13431	13498
130000	13712	13805
135000	13907	14106
140000	14183	14402
145000	14548	14692
150000	14911	14977
155000	15204	15257
160000	15423	15533
165000	15656	15804
170000	15973	16071
175000	16232	16334
180000	16481	16594
185000	16692	16849
190000	16980	17101
195000	17218	17349
200000	17454	17594
205000	17741	17836
210000	17985	18074
215000	18284	18310
220000	18434	18543
225000	18585	18773
230000	18777	19000

235000	19062	19224
240000	19360	19447
245000	19614	19666
250000	19885	19883
255000	20058	20098
256560	20094	20165

Extrapolated Tokens	Expected Types
500000	28653
750000	35118
1000000	40383

Table 7.4 Calculation and extrapolation of growth of types in the English corpus

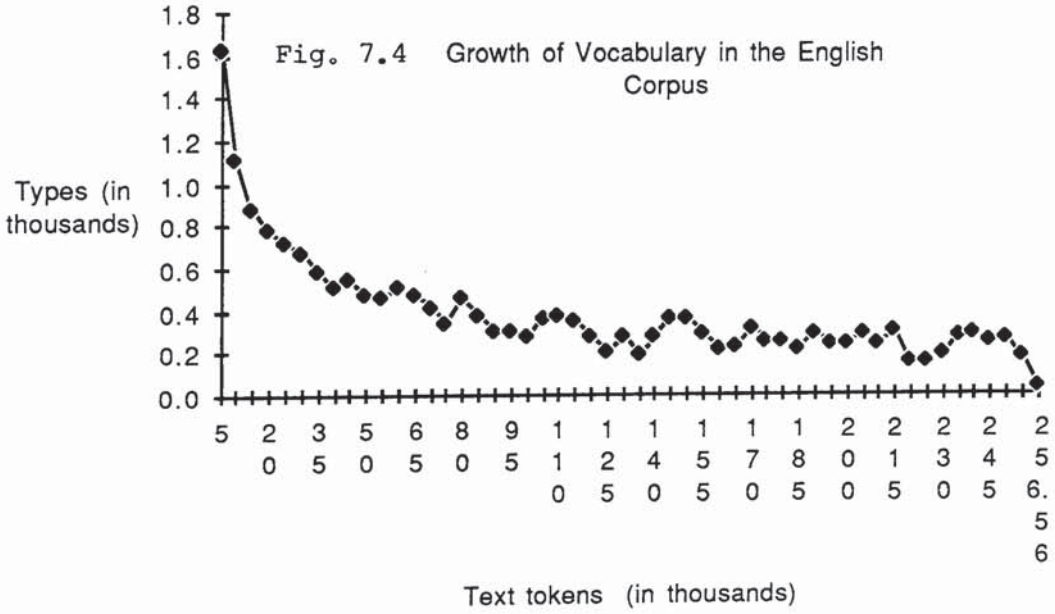


Fig. 7.4 Growth of Vocabulary in the English Corpus

Examination of the growth of types show that the first two intervals contain, as would be expected, the highest number of types. Growth starts to slow down gradually. At one interval (interval 44) there is a considerable degree of saturation of vocabulary and growth is, therefore, at its lowest: only 150 new types.

Extrapolation to a corpus of 500,000 words is expected to contain 28,653 types. An extrapolated corpus of a million tokens is expected to yield 40,383 types. This figure is noticeably smaller than the observed number of types in the Brown and LOB corpora (which is more than 50,000). This looks rather odd considering the fact that the double logarithmic type-token ratio is almost identical in all three corpora. However, one can argue on the basis of Tuldava's indices of richness and growth rate that the manner of growth of types is, to some extent, different statistically in our corpus as compared to the other two corpora. In general this requires deeper analysis and the use of more varied and more sophisticated statistical procedures to investigate the quantitative nature of our corpus in comparison to the LOB and Brown, a task that goes beyond the scope of this work and may be reserved for future work.

7.3.2 Statistical Profile of the Arabic corpus

7.3.2.1 Frequency and Rank Distributions

As stated earlier (Ch. 5) the computation of the Arabic corpus is not based on the orthographic word, but on our own conception of what constitutes "word" as a unit of linguistic measurement (see a detailed discussion in App. 1). We believe this conception of the

word in Arabic is a valid means for computing a text corpus, though, admittedly, it is not the only one. For instance, one can perform a series of calculations on the basis of the lemmatised forms or even the roots. Or, alternatively, one can use ad hoc procedures whereby certain grammatical words are deleted, particularly those that are orthographically connected to other words preceding or following them. However, each of these procedures reflects different theorisation or is intended for a different set of aims and may, therefore, be convenient for a different type of project.

According to the statistics that we have obtained from the OCP runs, the Arabic corpus comprises 256,476 word tokens. The number of types is 16,434, considerably lower than the number of types in the English corpus. The number of hapaxes is 7,173.

A SPITBOL program that is designed to ignore some dialectal expressions in the corpus while measuring types and token gives slightly different results. The text length is 256,450 word tokens and the size of the vocabulary is 16,402 types. These results are used in other calculations in the corpus.

The rank distribution consists of 252 ranks, a lower figure than its English counterpart. The first top ten ranks, occupied by one type each, achieve 38.85% of token coverage, though not more than 0.06% of total types. This is a wider coverage than the one achieved by the top ten ranks of the English corpus. Indeed a coverage of this size is achieved in the English rank distribution by the first top 38 ranks. This is an interesting point and reflects an important role of the most frequent words, all grammatical (synsemantic or function) words, in discourse. The next top ten ranks, again occupied by 1 type each, represents only 6.12%

of the total number of tokens, a figure that tallies to a large extent with its counterpart in the English rank distribution. Accordingly, the twenty most frequent types achieve about 45% of token coverage.

7.3.2.2 Type-Token Measurements

The type-token ratio in the Arabic corpus is 0.064 considerably lower than the same figure in the English corpus. The logarithmic type-token ratio is 0.779, and the double logarithmic type-token ratio is 0.852, both figures are still lower than their counterparts in English. Figure (7.5) plots the frequency of types against tokens, and Figure (7.6) contrasts the type-token distributions in both corpora.

The calculations of measurements based on the relations between the size of the vocabulary, the text length and the hapax legomena are shown in Table (7.5). The concentration index is 0.036 lower than its English counterpart. This is due to low vocabulary size and to a lower number of hapaxes. The hapax/non-hapax ratio in the English corpus is 0.9070, while it is lower in the Arabic corpus: 0.7745. This fact affects the rest of the calculations. Exclusivity index is 0.028, lower than its English counterpart. The hapax probability is 0.435, again, lower than the hapax probability in the English corpus. The non-hapax probability is 0.565, slightly higher than its corresponding figure in the English corpus. The index of variegation is 43.5% indicating lower diversification of vocabulary in Arabic than in English. The gravity index, which represents the rate of hapax occurrences is 35.93, substantially higher than the corresponding figure in the English corpus.

Concentration:	0.0361337
Consolidation factor:	0.564923
Density:	18.4827
Exclusivity:	0.0278285
Gravity:	35.9344
Hapax probability:	0.435077
Non-hapax probability:	0.564923
Lexical efficiency:	0.636059
Lexical usage:	0.495645
Type-token ratio:	0.0639622
Predictability:	0.936038
Rhythmicality:	25.9048
Stereotypicality:	26.9048
Type occurrence rate:	15.6342
Variegation:	43.5077

Table 7.5 Indices Based on Type and Token measurements in the Arabic Corpus

Fig. 7.5 Distribution of types and tokens in the Arabic Corpus

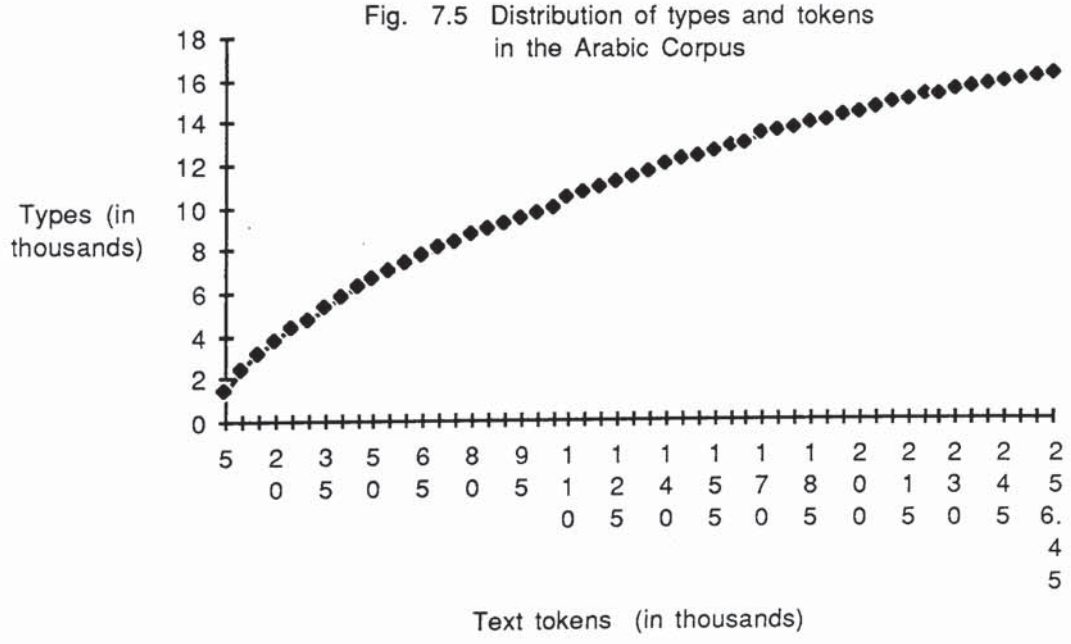
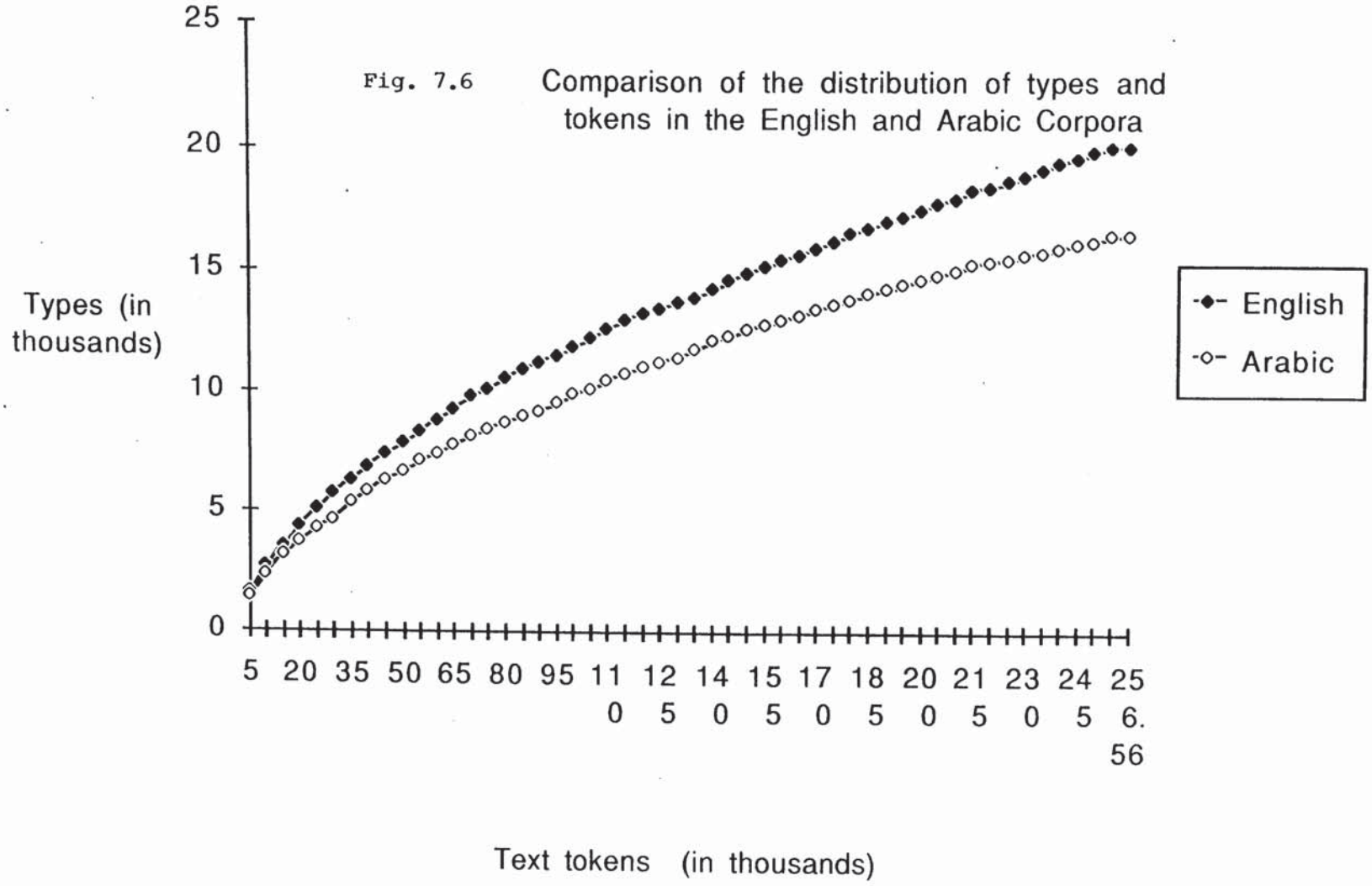


Fig. 7.6 Comparison of the distribution of types and tokens in the English and Arabic Corpora



The type occurrence rate is 15.6 giving the Arabic corpus a larger text span than the case in the English corpus before a new type is encountered. The index of stereotypicality is 26.9, larger than the corresponding figure in English and indicating more vocabulary being stereotypical, that is more repeatedness, since this figure indicates the "mean" of number of repeated occurrences for a non-hapax. The rhythmicality index is 25.9, again higher than the corresponding index in the English corpus. It indicates, again, a bigger size of repeatedness for the vocabulary. The intensity of the repeatedness is calculated at 0.525, which is the consolidation factor. Vocabulary density is 18.4827, lower than in English. Lexical efficiency is computed at 0.636, and the index of lexical usage is 0.49565, both lower than their counterparts in English.

7.3.2.3 Sentence Length

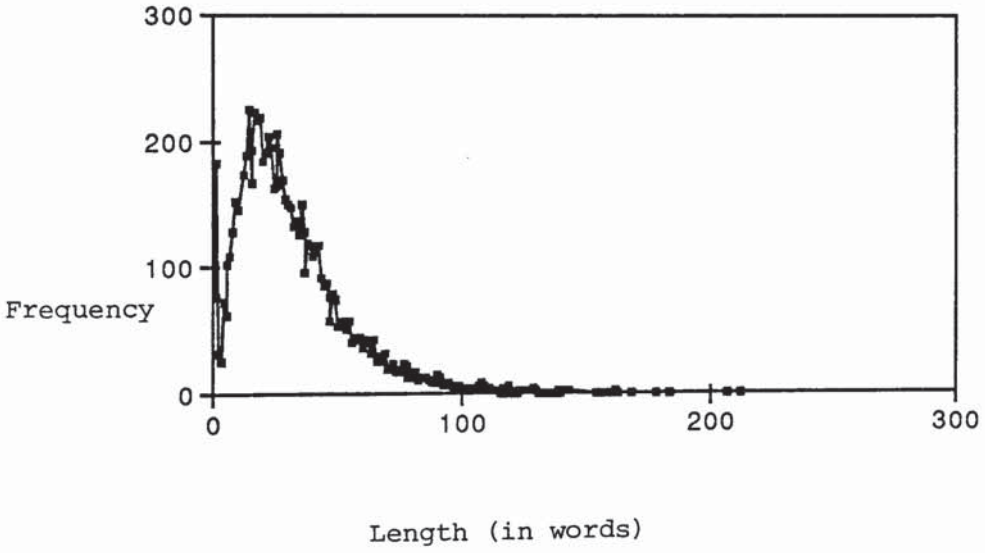
The results of the calculation of sentence length in the Arabic corpus are displayed in Table (7.6) (see also Appendix 11). The total number of sentences found in the corpus are 8060, considerably lower than the number of sentences in the English corpus. The average sentence length is 31.83 words, which is bigger than the average length of the English sentence. The measure of confidence shows that in 95% of cases the average sentence length ranges between 31.3453 and 32.3178. The standard deviation of sentence length is 22.2737, larger than the corresponding figure in English.

The longest sentence, occurring only once, comprises 212 words. The shortest sentence consists of only one word, which occurs 77 times. The most frequent sentence length in Arabic is, surprisingly, rather low: 14 words occurring 224 times, i.e. 2.78%

Total number of words	256562
Total number of sentences	8060
Average sentence length	31.8316
Variance	496.0113
Standard deviation	22.2737
Coefficient of variation	69.9735%
Standard error	0.2481
Lower 95% confidence limit	31.3453
Upper 95% confidence limit	32.3178

Table 7.6 Sentence length in the Arabic corpus

7.7 Distribution of Sentence Length in the Arabic Corpus



of total sentences. This sentence length is nearly similar to its English counterpart, which is 15 words, though the latter length is comparatively more frequent. The coefficient of variation of sentence lengths in Arabic is high, 70%. Figure (7.7) shows the frequency distribution of sentence length in the Arabic corpus.

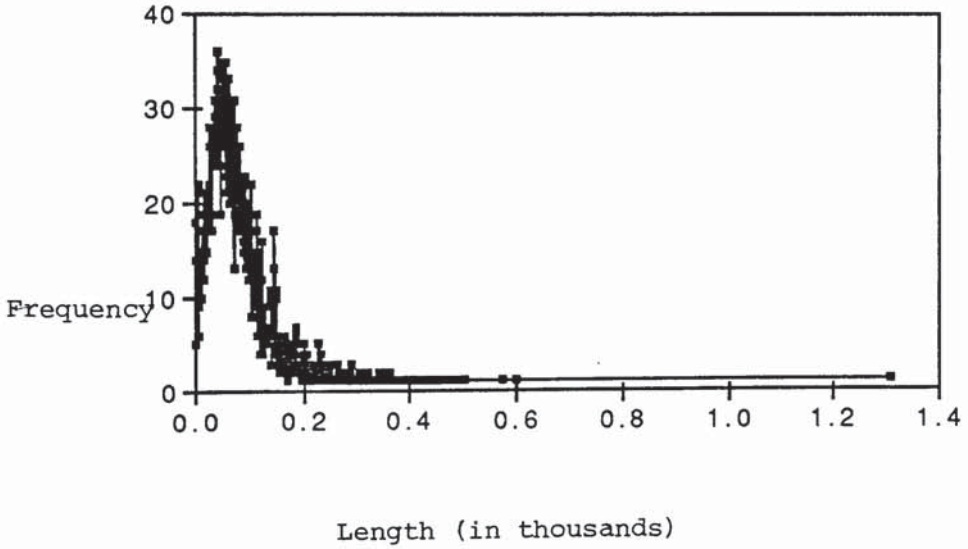
7.3.2.4 Paragraph Length

The measurement of the number and length of paragraphs in Arabic offers insights into the way text is structured in Arabic as compared to English (see Ch. 8 for a discussion). The result of the calculation are displayed in Table (7.7) and a fuller description is given in Appendix (13). The Arabic corpus comprises 3033 paragraphs, a considerably lower figure than its counterpart in English. This demonstrates that the average paragraph in Arabic is longer in terms of words than in English, the length being 84.6 words. Furthermore, the standard deviation, calculated at more than 72, indicates a substantially great departure from the mean length of the paragraph. For instance, the longest paragraph, which has 1 occurrence, is 1306 words while the shortest one, occurring 5 times, comprises only 1 word. A third point concerns the wide variations in the distribution of lengths. This is stipulated in the coefficient of variation, which is calculated at 85%. The most frequent paragraph length is 39 words with a frequency of 36 making up approximately 1.2% of the total number of paragraphs. Figure (7.8) gives a visual description of the distribution of paragraph lengths in the Arabic corpus.

Total number of paragraphs:	3033
Total number of words:	256562
The smallest paragraph had	1 words
The largest paragraph had	1306 words
The average paragraph length is	84.59 words
Variance is	5196.00092
Standard deviation is	72.00896
Coefficient of variation is	85.22%

Table 7.7 Paragraph length in the Arabic corpus

Fig. 7.8 Distribution of Paragraph Lengths in the Arabic Corpus



7.3.2.5 Calculation of Growth

Measurement of growth within intervals of 5000 words of running text indicates a smaller growth of vocabulary in Arabic than in English. The results are given in Table (7.8) and displayed in Figure (7.9). Comparison of the growth of vocabulary in both corpora is illustrated in Figure (7.10). The first interval contains 1,447 types, lower than its corresponding figure in English. This lower growth figures are consistent within most intervals. The lowest growth is in interval 45 where the observed number of new types is 111. Tuldava's extrapolation procedure shows that the expected vocabulary size in a corpus of 500,000 words is 23,495. An extrapolated corpus of a million words is expected to contain 32,890 types, a significantly smaller vocabulary size than its counterpart in English.

Fig. 7.9 Growth of Vocabulary in the Arabic Corpus

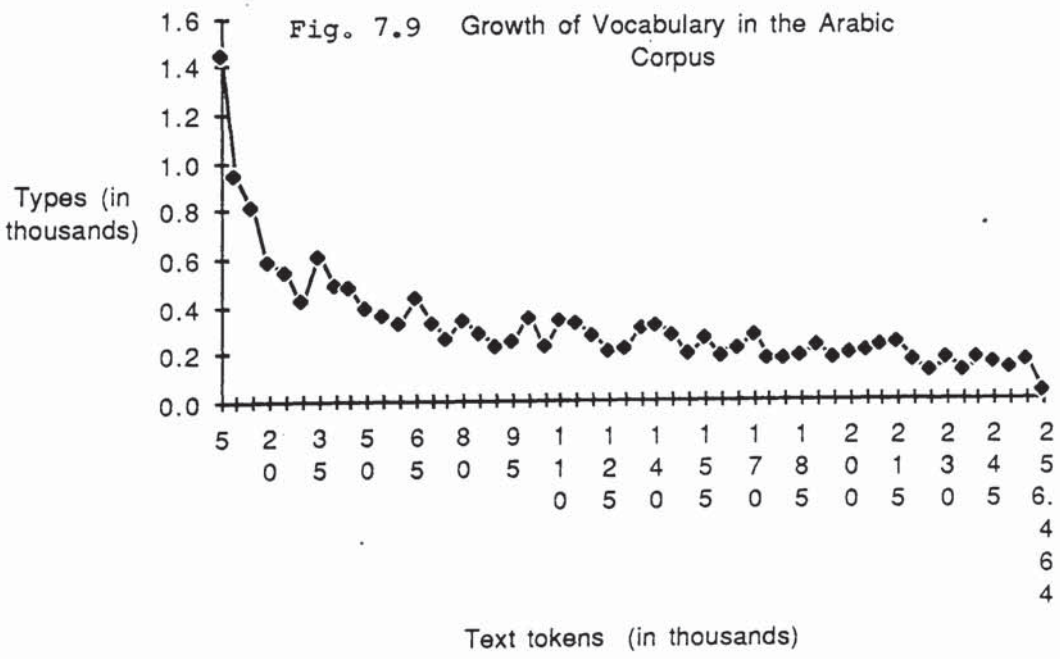
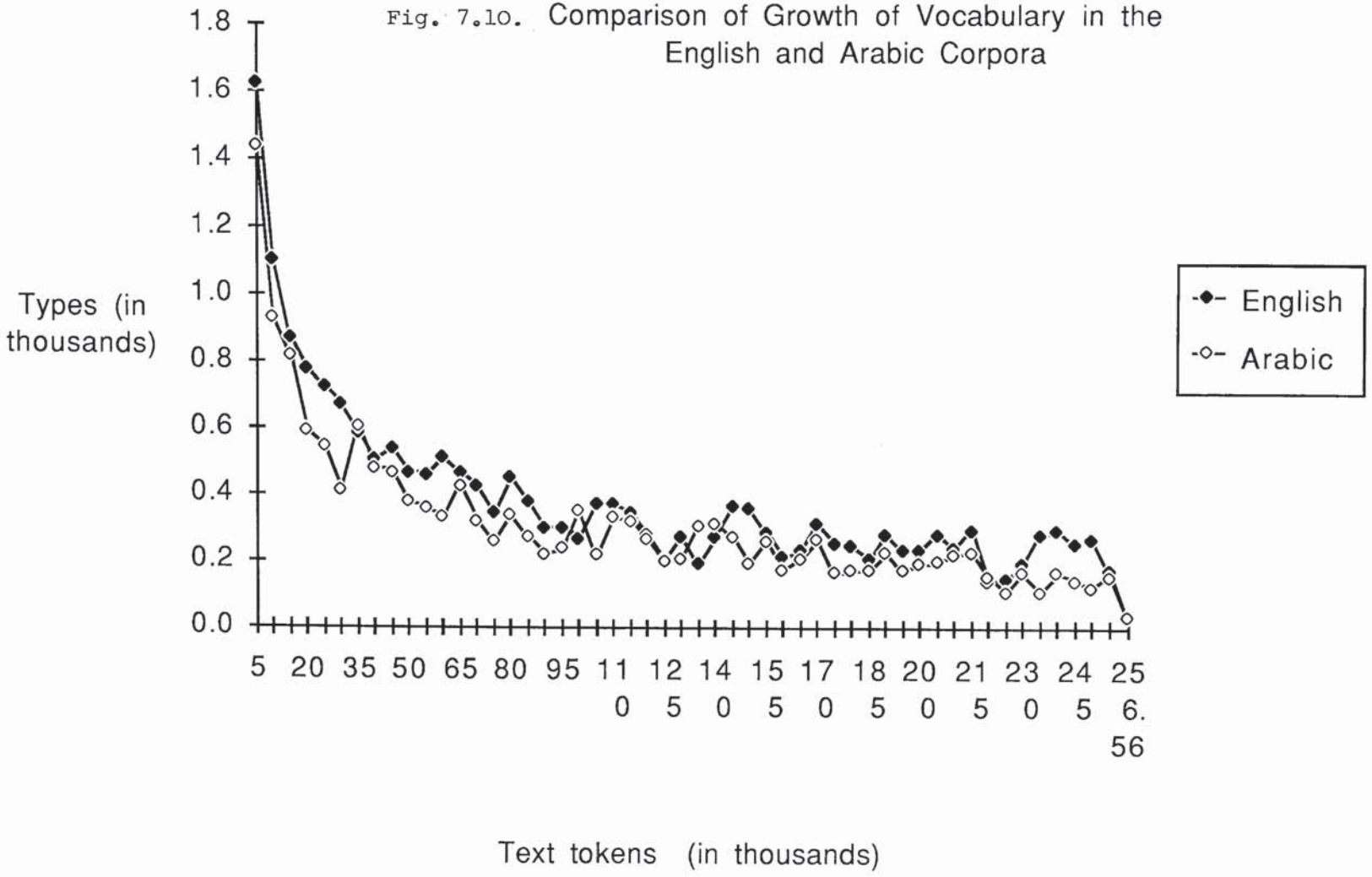


Fig. 7.10. Comparison of Growth of Vocabulary in the English and Arabic Corpora



It should be noted that more discussion based on global measurements of the corpora is possible but goes beyond the scope of this chapter. Some contrastive statements based on the calculations that we have made will be given in the next chapter. Meanwhile, we shall discuss results of the computation of connectives in the two corpora.

Tuldava's richness index: 0.012261
Vocabulary growth rate: 0.010718

Actual Tokens	Actual Types	Expected Types
5000	1447	1490
10000	2383	2389
15000	3204	3120
20000	3798	3755
25000	4346	4325
30000	4764	4847
35000	5373	5331
40000	5858	5785
45000	6332	6213
50000	6711	6620
55000	7070	7007
60000	7402	7379
65000	7831	7736
70000	8160	8079
75000	8421	8411
80000	8764	8733
85000	9042	9044
90000	9267	9347
95000	9510	9641
100000	9865	9927
105000	10085	10207
110000	10421	10479
115000	10750	10746
120000	11018	11007
125000	11221	11262
130000	11433	11511
135000	11742	11756
140000	12061	11997
145000	12336	12232
150000	12530	12464
155000	12791	12692
160000	12967	12915

165000	13178	13135
170000	13451	13352
175000	13620	13565
180000	13791	13775
185000	13967	13982
190000	14195	14186
195000	14366	14387
200000	14560	14585
205000	14760	14781
210000	14983	14974
215000	15215	15165
220000	15374	15353
225000	15485	15539
230000	15652	15722
235000	15765	15904
240000	15931	16083
245000	16080	16261
250000	16205	16436
255000	16365	16609
256464	16402	16660

Extrapolated Tokens	Expected Types
500000	23495
750000	28678
1000000	32890

Table 7.8 Calculation of growth of vocabulary in the Arabic Corpus

7.4 Quantitative Characteristics of Connectives: A General Profile

7.4.1 Preliminaries

Before we proceed into making a description of the quantitative characteristics of connectives, an introductory note is in order. This statistical profile has two main features:

1. It is a general account of connectives as forms, i.e. connectives are not differentiated at this stage on this basis of functionality. Hence a connective such as "and" is considered one type regardless of the various textual functions it serves in the text.

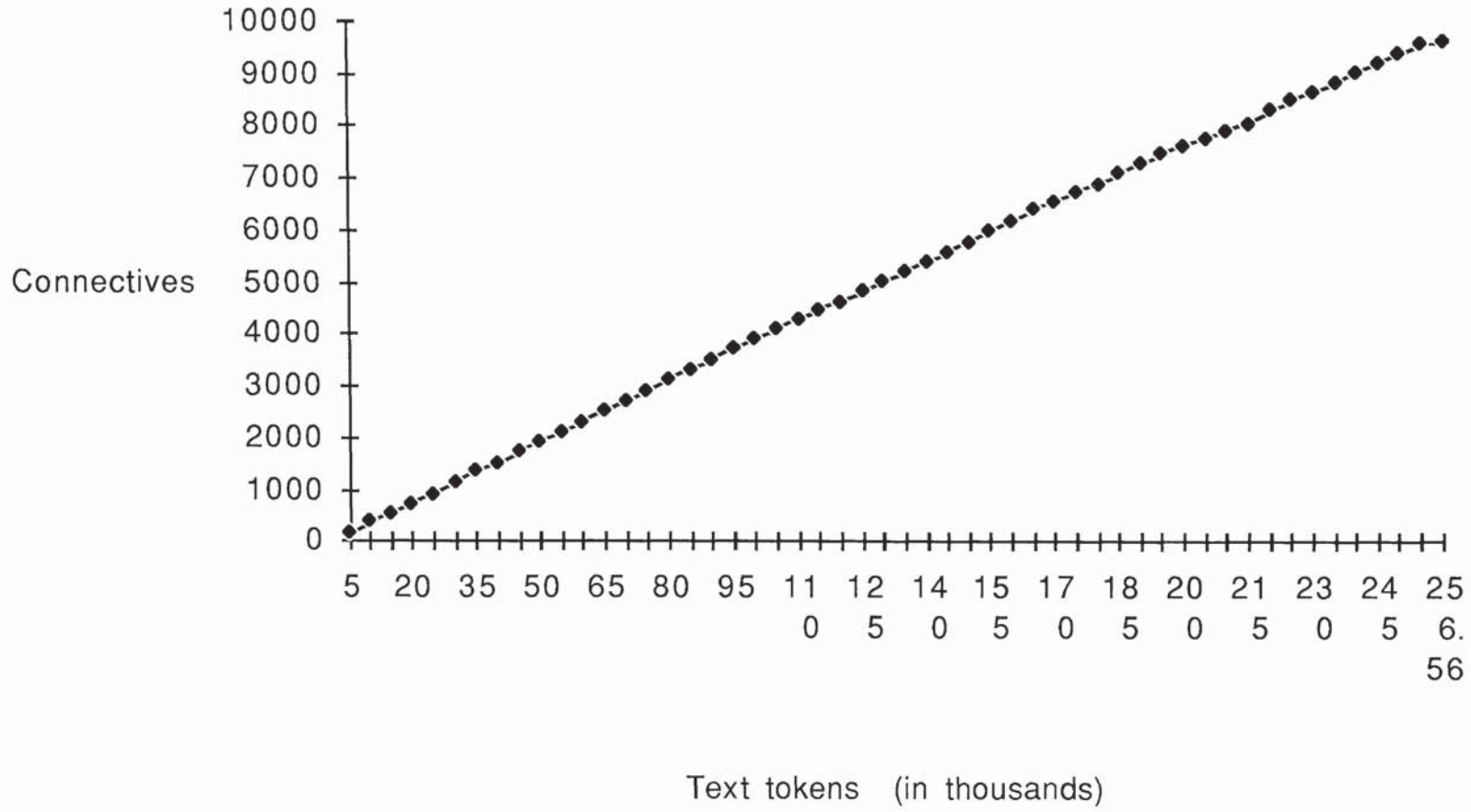
2. Connectives, as both types and tokens, refer in this profile to connective "cores", i.e. the central word-elements in connective structures. As discussed previously (see Ch. 5), a connective has a core that bears the semantic meaning, and that represents the key word under which a connective usually appears in the dictionary. This core also represents the element that carries the tag in the corpus. A one-word (or simple) connective constitutes a core, which is the connective form itself, e.g. "and", "but", etc., but no peripherals. A multi-word connective (compound or correlate) has a central element and one or more peripherals, e.g. "on the one hand", "either ... or". In this case, all peripherals are regarded as part of the text for the purposes of the measurements; only "cores" are included in the calculations. However, there are instances in this profile where calculations related to compound connectives as full units are indicated for the purpose of clarity or comparison. All cores of multi-word connectives are listed within Appendices (14-5).

7.4.2 General Profile of English Connectives

7.4.2.1 Frequency and Rank Distribution

The total number of connective tokens observed in the English corpus is 9,596 and the number of connective types (computed as connective cores) is 311. The complete frequency lists of connectives are to be found in Appendices (14 and 16); the first list is sorted alphabetically while the second according to descending frequency. A full concordance of English connectives,

Fig. 7.11 Distribution of Connectives in the English Corpus



with key words arranged alphabetically, is given in the microfiche appendices. Figure (7.11) displays the distribution of connective tokens in the corpus.

The influence of high frequency connectives can be seen in the distribution given in Table (7.9) and displayed in Figure (7.12).

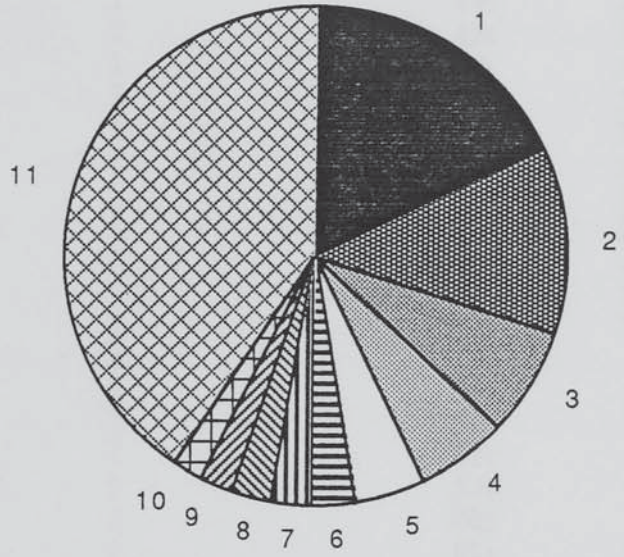
Rank	Connective	Abs. Freq	Rel. Freq		Cum. Freq
1:	and - - - -	1721	17.9346%	1721	17.9346%
2:	but - - - -	1180	12.2968%	2901	30.2314%
3:	if - - - - -	669	6.9717%	3570	37.2031%
4:	as - - - - -	555	5.7837%	4125	42.9867%
5:	when - - - -	415	4.3247%	4540	47.3114%
6:	also - - - -	290	3.0221%	4830	50.3335%
7:	because - -	247	2.5740%	5077	52.9075%
8:	even - - - -	228	2.3760%	5305	55.2835%
9:	then - - - -	221	2.3030%	5526	57.5866%
10:	so - - - - -	199	2.0738%	5725	59.6603%
11:	however - -	190	1.9800%	5915	61.6403%
12:	while - - -	151	1.5736%	6066	63.2139%
13:	or - - - - -	128	1.3339%	6194	64.5478%
14:	though - - -	125	1.3026%	6319	65.8504%
15:	since - - -	116	1.2088%	6435	67.0593%
16:	indeed - - -	108	1.1255%	6543	68.1847%
17:	than - - - -	105	1.0942%	6648	69.2789%
18:	of course -	102	1.0629%	6750	70.3419%
19:	although - -	101	1.0525%	6851	71.3944%
20:	yet - - - -	100	1.0421%	6951	72.4365%
21:	for - - - -	97	1.0108%	7048	73.4473%
22:	again - - -	76	0.7920%	7124	74.2393%
23:	before - - -	74	0.7712%	7198	75.0105%

Table 7.9 Connectives achieving 75% token coverage in the English corpus

The 10 most frequent connectives achieve approximately 60% of token coverage. These are and, but, if, as, when, also, because, even, then, and so. Another 13 connectives, in addition (i.e. total 23 connectives), achieve 75% token coverage. These are however, while,

or, though, since, indeed, than, of course, although, yet, for, again, before. The total number of connective types that achieve 90% token coverage is 68 (i.e. 45 types represent about 15% of total tokens). The last 10% of token coverage is achieved by 243 types.

Fig. 7.12 Distribution of the 10 Most Frequent Connectives in the English Corpus



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

This situation can be explained by reference to the rank distribution of connectives (see Table 7.10). There are 63 ranks in the distribution. The top rank is occupied by one connective that has 1,721 occurrences. This is the connective "and". This rank represents 18% of total token occurrences. The next rank is occupied by the connective "but" with a frequency of 1,180 representing 12.3% of total connective tokens. The first 31 ranks in the distribution are occupied by one connective each.

Rank	Tokens	Types	Rank	Tokens	Types
1:	1721	1	33:	40	1
2:	1180	1	34:	39	1
3:	669	1	35:	36	1
4:	555	1	36:	34	1
5:	415	1	37:	33	3
6:	290	1	38:	32	2
7:	247	1	39:	31	3
8:	228	1	40:	29	1
9:	221	1	41:	27	1
10:	199	1	42:	25	2
11:	190	1	43:	23	2
12:	151	1	44:	22	5
13:	128	1	45:	20	4
14:	125	1	46:	18	1
15:	116	1	47:	17	3
16:	108	1	48:	16	2
17:	105	1	49:	15	4
18:	102	1	50:	14	3
19:	101	1	51:	13	1
20:	100	1	52:	12	3
21:	97	1	53:	11	7
22:	76	1	54:	10	9
23:	74	1	55:	9	5
24:	69	1	56:	8	8
25:	66	1	57:	7	13
26:	65	1	58:	6	14
27:	63	1	59:	5	20
28:	61	1	60:	4	15
29:	58	1	61:	3	23
30:	54	1	62:	2	35
31:	46	1	63:	1	85
32:	41	2			

Number of connective tokens in the English corpus: 9596
Number of connective types in the English Corpus: 311

Table 7.10 Rank distribution of connectives in the English Corpus

The lowest rank in the distribution is filled by a large number of hapax legomena. These comprise 85 connectives (see Appendix 16). Although the hapaxes make up only 0.89% of total connective tokens, they constitute 27.3% of total types.

The lowest rank of the non-hapaxes is occupied by 35 connectives

each with two occurrences. This makes up 0.73% of total connective tokens and represents a large portion of non-hapax types, nearly 15.5% (and 11.25% of total types).

7.4.2.2 Connective-sentence Distribution

The next measurement calculates the relationship between the distribution of connectives and the distribution of sentences. The results of the calculations are presented in Table (7.11) and Figure (7.13). The number of sentences that do not contain connectives are 5,391, representing 46% of total number of sentences (which is calculated at 11,671). The number of sentences that contain connectives is 6,280, making up 54% of the total number of sentences. Sentences that have one connective each total 3,983, a figure that represents about 34% of total sentences and 63.4% of the number of sentences containing connectives. Sentences that contain two connectives each total 1,579, making up 13.5% of the total number of sentences and 25% of the sentences containing connectives. The total number of sentences containing more than two connectives is 718, which represents only 6.4% of the total number of sentences and 11.43% of sentences having connectives. There are two occurrences of sentences with 8 connectives each and one occurrence of a sentence containing ten connectives.

The average connective per sentence in the English corpus is calculated at 0.8222. The confidence index reveals that in 95% of the cases the average connective per sentence ranges between 0.8044, the lower limit, to 0.8400, the upper limit. The standard deviation is 0.9825 and the coefficient of variation is 119%.

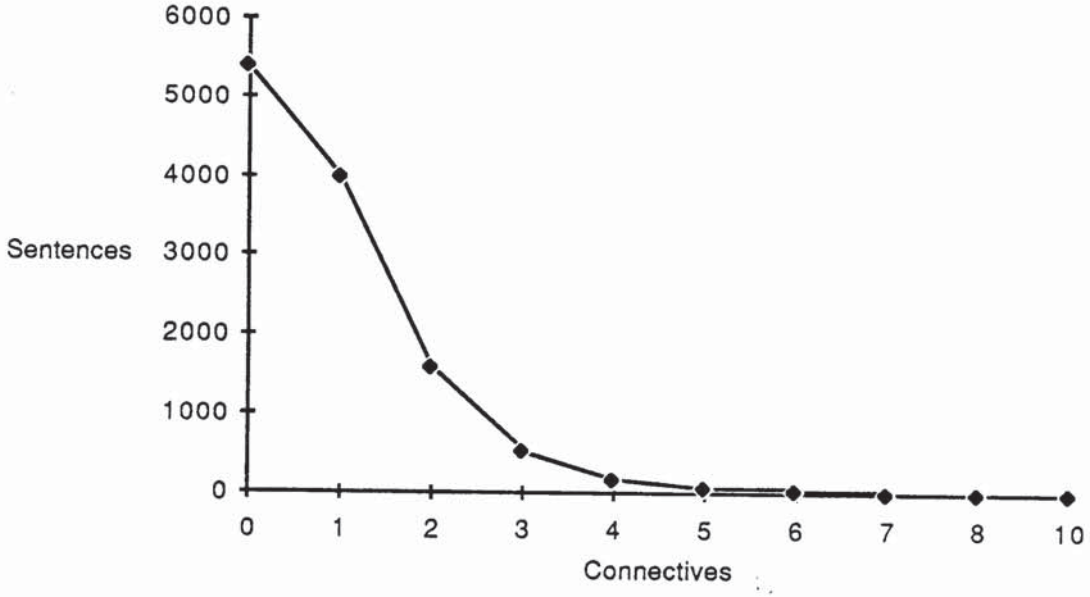
No of Conn	No of Sen	Cum Conn	Cum Sen	Rel Freq	Cum Freq
0:	5391	< 0>	[5391]	46.191%	[46.191%]
1:	3983	< 3983>	[9374]	34.127%	[80.319%]
2:	1579	< 7141>	[10953]	13.529%	[93.848%]
3:	509	< 8668>	[11462]	4.361%	[98.209%]
4:	148	< 9260>	[11610]	1.268%	[99.477%]
5:	43	< 9475>	[11653]	0.368%	[99.846%]
6:	10	< 9535>	[11663]	0.086%	[99.932%]
7:	5	< 9570>	[11668]	0.043%	[99.974%]
8:	2	< 9586>	[11670]	0.017%	[99.991%]
10:	1	< 9596>	[11671]	0.009%	[100.000%]

Total number of connectives	9596
Total number of sentences	11671
Average connectives per sentence	0.8222
Variance	0.9654
Standard deviation	0.9825
Coefficient of variation	119.0005%
Standard error	0.0091
Lower 95% confidence limit	0.8044
Upper 95% confidence limit	0.8400

Table 7.11 Distribution of Connectives per Sentence in the English Corpus

Code of columns:Conn = Number of connectives at any one sentence
Sen = Number of sentences having numbers of connectives specified in column Conn.
Cum Conn = Cumulation of connective tokens (result of Conn x Sen)
Cum Sen = Cumulation of sentences
Rel Freq = Relative Frequency of Figures in Sen
Cum Freq = Cumulation of Figures in Rel Freq

Fig. 7.13 Distribution of Connectives per Sentence in the English Corpus



7.4.2.3 Type-token measurements

Results of type-token measurements are summarised in Table (7.12). Connectives in the English corpus have a type-token ratio of 0.0324. The logarithmic TTR is 0.6261 and the double logarithmic TTR is 0.6611. These indices represent richness and extensity of connective types as related to tokens in the corpus. Figure (7.14) plots the types against the tokens in the corpus. The number of hapax legomena is 85 and the number of non-hapaxes is 226.

Concentration:	0.0235515
Consolidation factor:	0.726688
Density:	13.3464
Exclusivity:	0.00885786
Gravity:	112.894
Hapax probability:	0.273312
Non-hapax probability:	0.726688
Lexical efficiency:	0.65669
Lexical usage:	0.411083
Type-token ratio:	0.0324093
Predictability:	0.967591
Rhythmicality:	41.0841
Stereotypicality:	42.0841
Type occurrence rate:	30.8553
Variegation:	27.3312

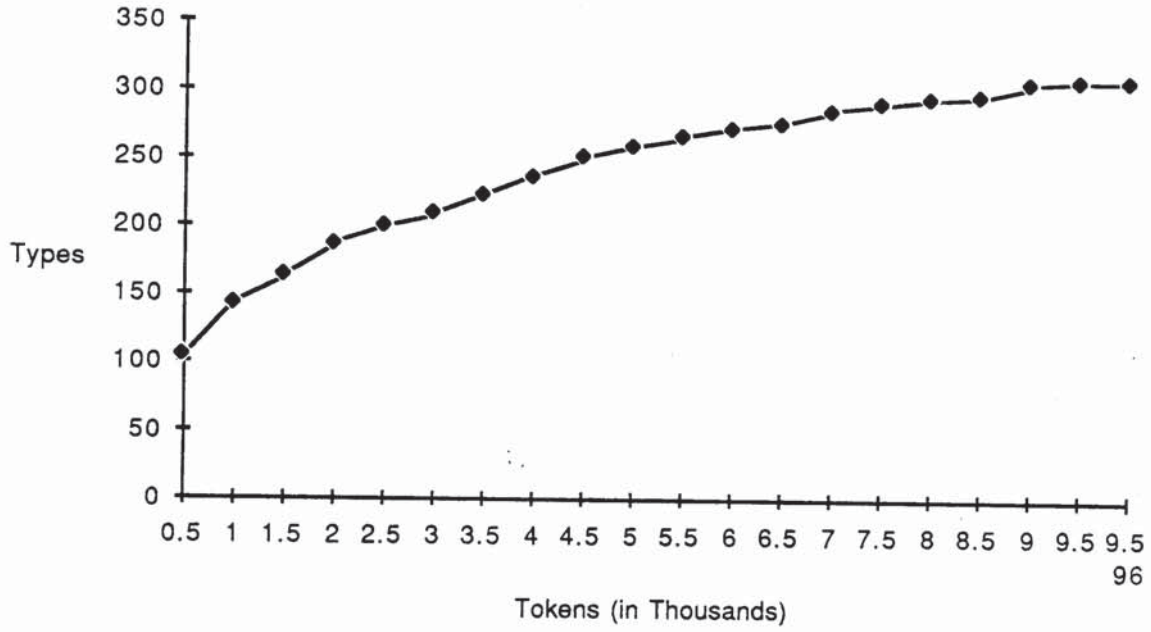
Table 7.12 Indices based on connective type-token measurements in the English corpus

The concentration index is calculated at 0.02355. This is related to non-hapaxes and represents their extensity of use in relation to connective token mass and can be taken as an index of the extensity of repeatedness. The exclusivity index, representing the size of the share that hapaxes reserve in token mass, is 0.008858. The variegation index, indicating the size of hapaxes in relation to the types, and therefore representing connective diversification, is 27.3312%. The hapax probability of occurrence is 0.273 while the non-hapax probability is 0.727. The gravity index, indicating the rate of occurrence of hapaxes, is 112.765.

The type occurrence rate of connectives is about 31. This index stipulates the number of connective tokens that we may have before we encounter a new type. The index of density is calculated at 13.346. The stereotypicality index, quantifying the repeatedness of

non-hapaxes, is computed at 42.084. The rhythmicality measure is 41.084. The intensity of repeatedness is calculated at 0.727, which is the consolidation factor. The predictability index is 0.968.

Fig. 7.14 Distribution of Connective Types and Tokens in the English Corpus



7.4.2.4 Entropy and redundancy

The rank distribution is used for working out the entropy and redundancy indices. The full rank calculation of entropy is given in App. (18). The maximum possible entropy for a set of 9,596 connectives is 13.2282 bits. This represents the calculation of H', indicating an entropy of an equidistribution of connectives. The entropy H is calculated at 5.4379 bits. This figure indicates the index of unexpectedness or uncertainty of English connectives. The

relative entropy, expressed as a percentage, is 41.1083%, while redundancy, which represents the presence of linguistic structuring, is 58.8917%.

7.4.2.5 Repeat Rate Indices

Three types of repeat rate are calculated for connectives.

1. The general repeat rate of connectives indicating the probability that two random selections from the whole text corpus will yield two connectives. The index is (14×10^{-4}), i.e. 14 in 10,000.

2. The connective system repeat rate indicating the probability that two random selections from the whole text corpus will give the same connective. The index is (88×10^{-6}), i.e. 88 in a million.

3. The connective type repeat rate indicating the probability that two random selections from the mass of connectives will yield the same connective type. The index is (63×10^{-3}), i.e. 63 in a 1,000.

A full list of the repeat rate of each connective type is given in Appendices (20-21).

7.4.2.6 Measures of Interval

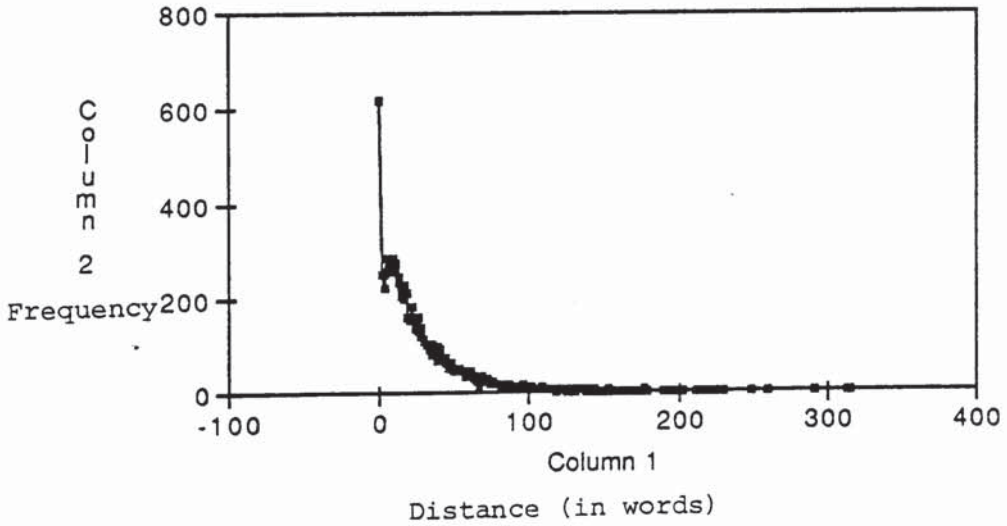
7.4.2.6.1 Gap Distribution

The first measure of interval is the distribution of gaps measured in terms of words. The results are given in Appendix (24). Note that in this measure we have considered as a gap the distance between sites occupied by connectives. In addition we have regarded as a gap the distance between the first word and the first

occurrence of a connective, and the distance between the site occupied by the last connective in the corpus and the end of the text. Figure (7.15) gives a visual representation of gaps between English connectives.

According to the distribution, there are 182 types of distance lengths. The most frequent one is 0 distance with a frequency of 618. This refers to the instances where two sites occupied by connectives are next to each other, i.e. with nonintervening words. Distance length of one word has 253 occurrences, which is the same figure for occurrences of two word distance lengths. The longest distance length observed is 315 words with only one occurrence. The average distance length is 25.7 words.

Fig. 7.15 Distribution of distance between occurrences of connectives in English



7.4.2.6.2 Levin's Measure of Interval

The next measurement that we applied is the calculation of Levin's index. The aim is to characterise the distribution of intervals that separate connectives in the corpus. In this respect, we consider the corpus as a continuous occurrences of a symbols (referring to connective tokens; type and function are immaterial in this measure) and B symbols (referring to all sites occupied by words other than connective tokens). Results of the calculation show that for a symbol occurring 9,596 times in the corpus, the L-compactness factor is 0.04967 and the Q-compactness factor is 0.00012. These measures indicate that the pattern of the repetitiveness of connectives has a tendency towards compactness throughout the text. Further, the high compactness value results from characteristic clustering in the way connectives are repeated.

7.4.2.7 Measure of Growth

In calculating the growth of connectives in the corpus, two calculations are made. The first is "global" and intends to investigate the growth of connective tokens and types within fixed intervals calculated in terms of text tokens. The second is "local"; it computes the growth of connective types within intervals of connective tokens. The results are used for finding out the expected number of connectives within texts of given size lengths. The aim behind the two calculations is to establish the dependence of the number of connectives on the length of the text and on the number of tokens.

1. Global Growth of Connectives

Results of the first calculations are shown in Table (7.13). The corpus is divided into intervals of fixed length, each is set equal to 5,000 words of running text. Growth of connective tokens and types is computed within each interval. This is represented visually in Figure (7.16).

The first interval contains 173 connective tokens, which includes 62 connective types. According to Tuldava's calculation, the estimated number of tokens is 188, which is expected to comprise 70 types. The second interval contains 217 more connectives, which raises the number of tokens to 390. Growth of types start to drop slowly; the second interval contains only 29 new types, raising the number of types to 91. The expected figures are 190 tokens and 23 types.

Growth of connective tokens varies from 219 (at interval 10), which is its highest growth, to 142 (at interval 43), which is its lowest growth. This variation is related to chance fluctuations. But growth of connective types slows considerably as text size and connective tokens increase. This means that fewer connective types are located within each additional interval of text. Indication for saturation of types comes early in the corpus. For instance, there are only 6 new types at interval 7, 3 at interval 12, and only 2 at interval 17. The 50th type percentile occurs within interval 7 (35,000 words of running text); the 75th type percentile occurs within interval 20 (100,000 words); and the 90th type percentile

---

Tuldava's index for connective tokens richness 3.039861
Growth rate of connective tokens: -3.034197
Tuldava's index for connective type richness 0.034591
Growth rate of connective types: 2.000315

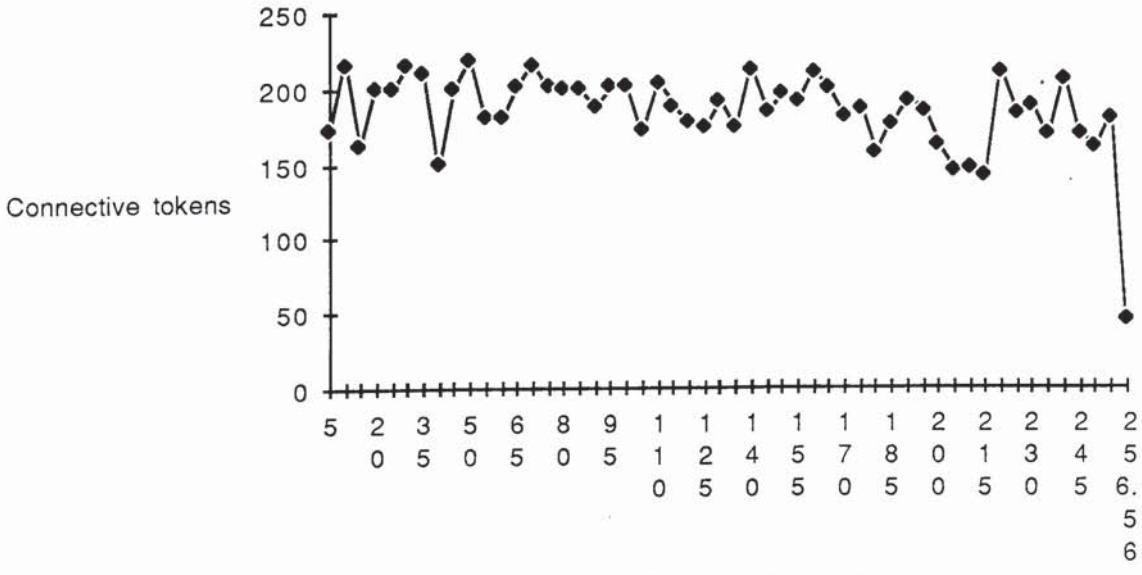
Actual Text Tokens	Actual Conn.Tokens	Expected Conn.Tokens	Actual Conn.Types	Expected Conn.Types
5000	173	188	62	70
10000	390	378	91	93
15000	552	569	108	109
20000	751	760	126	123
25000	950	952	142	134
30000	1167	1143	151	144
35000	1378	1335	157	153
40000	1529	1527	166	161
45000	1729	1719	175	169
50000	1948	1911	183	176
55000	2130	2103	190	183
60000	2311	2295	193	189
65000	2512	2488	200	195
70000	2728	2680	205	201
75000	2929	2872	209	206
80000	3128	3065	218	211
85000	3328	3258	220	216
90000	3517	3450	225	221
95000	3718	3643	228	225
100000	3920	3835	235	230
105000	4093	4028	241	234
110000	4296	4221	249	238
115000	4485	4414	253	243
120000	4663	4607	257	247
125000	4838	4800	260	250
130000	5031	4992	262	254
135000	5206	5185	263	258
140000	5419	5378	266	261
145000	5605	5571	269	265
150000	5802	5764	271	268
155000	5995	5958	274	272
160000	6206	6151	276	275
165000	6406	6344	278	278
170000	6588	6537	280	281
175000	6775	6730	282	284
180000	6932	6923	287	287
185000	7109	7116	289	290
190000	7301	7310	290	293
195000	7486	7503	293	296
200000	7648	7696	295	299
205000	7794	7890	296	302
210000	7941	8083	297	305
215000	8083	8276	297	307

220000	8294	8470	298	310
225000	8478	8663	299	313
230000	8667	8856	301	315
235000	8836	9050	305	318
240000	9042	9243	309	320
245000	9211	9437	310	323
250000	9371	9630	311	325
255000	9551	9824	311	328
256560	9596	9884	311	328

Extrapolated Text Tokens	Expected Connective Tokens	Expected Connective Types
500000	19318	420
750000	29027	488
1000000	38747	541
5000000	194916	960

Table 7.13 Calculation of "global" growth of connectives in the English corpus

Fig. 7.16 Growth of Connective tokens in the English Corpus



occurs within interval 34 (170,000 words). This early saturation of types after a fast growth is characteristic of English connectives.

However, saturation of connective types cannot be total. This is because connectives, as we have described them (see Chapters 4 and 6), are not a closed set of vocabulary items in the same way that grammatical (synsemantic) words are, and hence new types are expected to emerge as text size increases. But it is feasible, and probably necessary, to conceive of a maximum limit, though we cannot stipulate with any degree of confidence based on empirical evidence what that limit is and at what text length it is achieved. All we can do at this stage is to use extrapolation procedures to examine the expected growth if text size and connective tokens increase.

We, therefore, applied Tuldava's extrapolation formula on the set of numerical observations on the size of growth within the various intervals. In extrapolating for connective tokens, Tuldava's richness index is calculated at 3.003 and the growth rate of connective tokens in the corpus is -3.0286 (the negative symbol is created by using the double logarithm in the computation). An extrapolated text length of 500,000 word tokens gives an expected number of 19,181 connective tokens and 420 types. An extrapolated corpus of a million words is expected to contain 38,374 connective tokens and 541 types.

2. "Local" Growth

This measure computes the growth of connective types within intervals of connective tokens, regardless of the size of running text that each interval may represent. Connective tokens are

divided into sets of 500 successive connectives each. The corpus is scanned and once an end of an interval is reached, connective types are recognised, calculated and incremented to the figures of the previous interval. The results are exhibited in Table (7.14). Figure (7.17) gives a visual representation of growth within each interval.

The first 500 connective tokens contains 104 types giving the distinctly high ratio of 0.2. The next interval contains 40 new types, a relatively smaller number which reduces the ratio to 0.144. Increase in new types starts to drop at early intervals. For instance, at the end of interval 8 (4,000 connectives) the type-token ratio is reduced to 0.059 and at interval 15 (7,500 connectives) the ratio (0.039) starts to approximate the registered type-token of connectives (0.032). This is another evidence for a fast saturation of connective types in the English corpus.

In extrapolating for connective types, Tuldava's richness index is 0.041789 and the computed growth rate of types is 0.042229. An extrapolated number of 10,000 connective tokens has an expected number of 313 types, an increase of only 2 types over the observed number of types in an actual 9,500 connective tokens. An extrapolated number of 100,000 connective tokens gives an expected number of 451 types. This means that in addition to the list of observed connective types we can still expect to have another 140 types if text length is allowed to expand to the extent that it yields more than 10 times the number of observed tokens. One cannot help wondering what the nature of these types is, what adjustment, if any, they may impose on the functional categorisation of

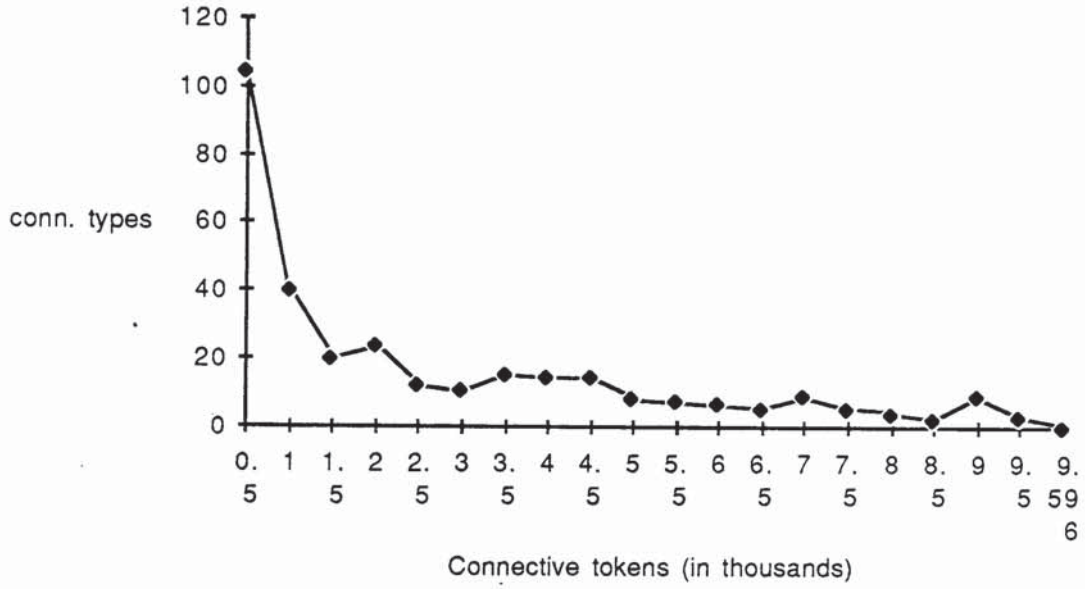
Tuldava's richness index:	0.041789
Connective growth rate:	0.042229

Actual Text Tokens	Actual Con.Tok	Actual Con.Types	Expected Con.Types
13519	500	104	102
26278	1000	144	141
39063	1500	164	168
51534	2000	188	188
64686	2500	200	204
76822	3000	210	218
89441	3500	225	229
102528	4000	239	240
115487	4500	253	249
129247	5000	261	257
142035	5500	268	265
155102	6000	274	272
167393	6500	279	278
182360	7000	288	284
195298	7500	293	290
212229	8000	297	295
225469	8500	299	300
238780	9000	308	304
253513	9500	311	309
256560	9596	311	309

Extrapolated Connective Tokens	Expected Connective Types
10000	313
15000	344
20000	365
30000	393
50000	423
100000	451

Table 7.14 Calculation of Growth of Connective types within intervals of connective tokens in the English corpus

Fig. 7.17 Growth of Connective Types within Tokens in the English Corpus



connectives, and how they may affect our understanding of the textual role of connectives. Probably only a massive and more variable text corpus may ever give us a definite answer (cf. discussion in Ch. 9).

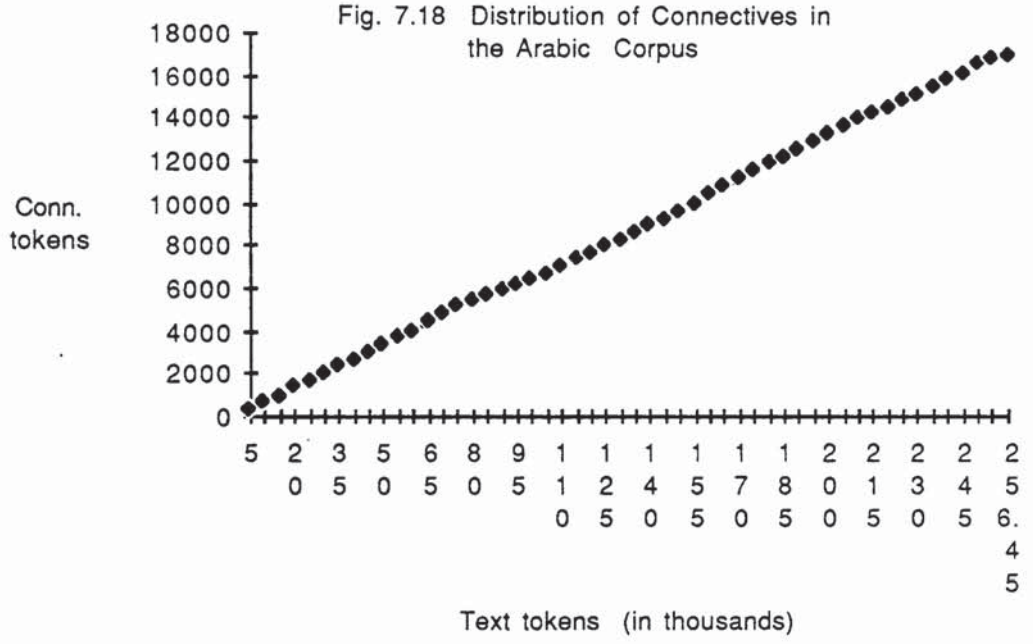
7.4.3 General Profile of Arabic Connectives

7.4.3.1 Frequency and Rank Distributions

In the Arabic corpus, the total observed number of connective tokens is 16,995, a figure that is 1.8 times the number of connective tokens in the English corpus. However, Arabic connective types (in terms of cores) is 297, slightly lower than its counterpart in English (which is 311 types). The complete frequency lists of connectives are given in Appendices (15 and 17), where entries are sorted alphabetically in the first and according to

descending frequency in the second. A full concordance of Arabic connectives is given in the microfiche appendices. Figure (7.18) plots the distribution of connective tokens within the corpus.

Fig. 7.18 Distribution of Connectives in the Arabic Corpus



High frequency connective types has a major influence on the overall distribution of connectives in Arabic. This is exemplified in Table (7.15). The 10 most frequent connectives in Arabic achieve 72.4% of token coverage. These are "wa, fa, kamā, iḏā, lākinna, li'anna, li, ḥattā, bal, 'aw". Figure (7.19) displays the shares in tokens that these connectives have in comparison to the rest of the types. The frequency of the next three connectives, "lākin", "ḏālika ('anna)" and "ayḍan", raises the token coverage percentage to more than 75%. The number of connective types that achieve 90%

of connective tokens is 45 (i.e. 32 connective types represent the next 15% of token coverage). The last 10% of tokens is achieved by 250 types.

The first most frequent connective, "wa", has a frequency of 8,208, making up 48.3% of total connective tokens. The second most frequent connective, "fa", has a frequency of 1879 which represents 11% of total connectives. The combined frequency of both these connectives achieve alone slightly less than 60% of connective

Rank	Connective	Abs. Freq.	Rel. Freq.		Cum. Freq.
1:	wa - - - -	8208	48.2966%	8208	48.2966%
2:	fa - - - -	1879	11.0562%	10087	59.3528%
3:	kamā - - -	372	2.1889%	10459	61.5417%
4:	'iḏā - - -	371	2.1830%	10830	63.7247%
5:	lākinna - -	311	1.0083%	11141	65.5546%
6:	li'anna - -	288	1.6946%	11429	67.2493%
7:	li - - - -	268	1.5769%	11697	68.8262%
8:	ḥattā - - -	216	1.2710%	11913	70.0972%
9:	bal - - - -	209	1.2298%	12122	71.3269%
10:	'aw - - - -	176	1.0356%	12298	72.3625%
11:	lākin - - -	173	1.0018%	12471	73.3805%
12:	ḏālika - -	170	1.0003%	12641	74.3808%
13:	'aẏḍan - -	161	0.9473%	12802	75.3281%

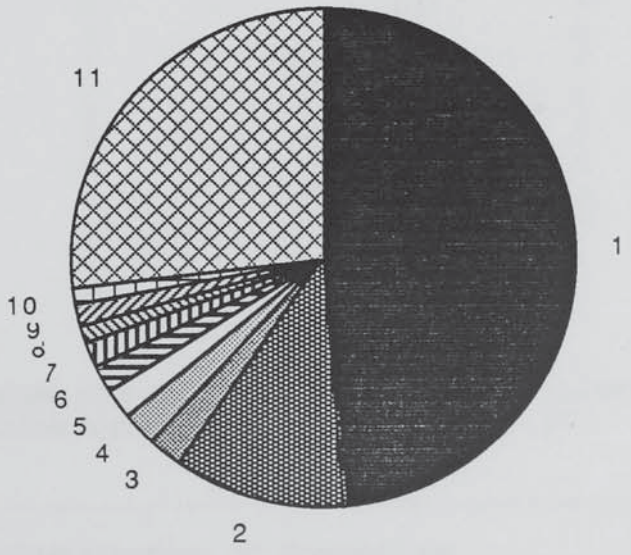
Table 7.15 Connectives achieving 75% of token coverage in the Arabic corpus

token coverage. This is an indication of the substantial reliance of textual connectivity in Arabic on the use of these two connectives.

The rank distribution of connectives is given in Table (7.16). There are 71 ranks. The first 30 ranks are occupied by one connective each. The lowest rank contains 77 hapaxes (see App. 17), making up about 26% of total types, though only 0.45% of total tokens.

The lowest rank of the non-hapaxes comprises 36 connectives each with two occurrences. This rank makes up 0.45% of total connective tokens but represents 16.4% of total non-hapaxes (and 12% of total types).

Fig. 7.19 Distribution of the 10 Most Frequent Connectives in the Arabic Corpus



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

Rank	Tokens	Types	Rank	Tokens	Types
1:	8208	1	37:	41	1
2:	1879	1	38:	40	3
3:	372	1	39:	38	1
4:	371	1	40:	37	2
5:	311	1	41:	34	1
6:	288	1	42:	33	1
7:	268	1	43:	31	2
8:	216	1	44:	29	2
9:	209	1	45:	28	1
10:	176	1	46:	27	3
11:	173	1	47:	25	2
12:	170	1	48:	24	1
13:	161	1	49:	23	1
14:	157	1	50:	22	3
15:	153	1	51:	21	3
16:	150	1	52:	20	2
17:	147	1	53:	19	3
18:	146	1	54:	18	4
19:	139	1	55:	17	2
20:	138	1	56:	16	2
21:	114	1	57:	15	6
22:	97	1	58:	14	3
23:	86	1	59:	13	4
24:	80	1	60:	12	5
25:	76	1	61:	11	4
26:	73	1	62:	10	8
27:	66	1	63:	9	7
28:	65	1	64:	8	6
29:	64	1	65:	7	9
30:	62	1	66:	6	6
31:	57	2	67:	5	8
32:	52	2	68:	4	12
33:	49	1	69:	3	27
34:	48	1	70:	2	36
35:	47	1	71:	1	77
36:	46	2			

Number of connective tokens in the Arabic corpus: 16995
Number of connective types in the Arabic Corpus: 297

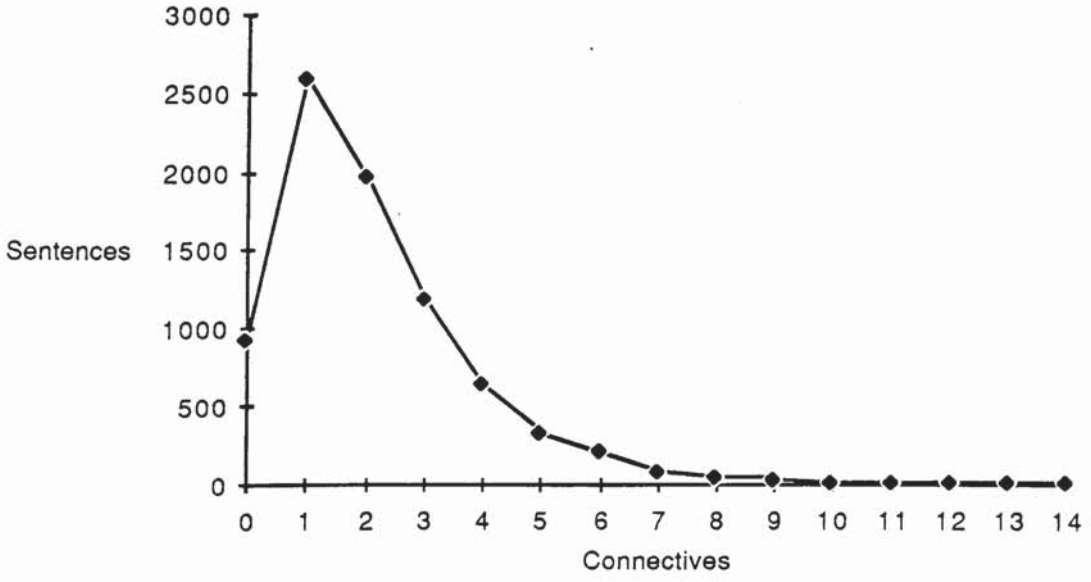
Table 7.16 Rank Distribution of Connectives in the Arabic Corpus

7.4.3.2 Connective-sentence distribution

The next measurement concerns the relationship between the distribution of connectives and the distribution of sentences. The results of the calculations are shown in Table (7.17) and displayed in Figure (7.20). The corpus comprises 8,060 sentences. Out of this figure only 933 sentences do not contain connectives, which represent 11.6% of the total. The number of sentences that contain connectives is 7,127, making up 88.4% of total sentences. These percentages are in sharp contrast with their counterparts in English and therefore form an important indicator of the role of connectives in Arabic text organisation.

Sentences having connectives are distributed in the following way. Sentences that contain one connective each make a total of

Fig. 7.20 Distribution of Connectives per Sentence in the Arabic Corpus



No of Conn	No of Sen	Cum Conn	Cum Sen	Rel Freq	Cum Freq
0:	933	< 0>	[933]	11.576%	[11.576%]
1:	2608	< 2608>	[3541]	32.357%	[43.933%]
2:	1983	< 6574>	[5524]	24.603%	[68.536%]
3:	1196	<10162>	[6720]	14.839%	[83.375%]
4:	633	<12694>	[7353]	7.854%	[91.228%]
5:	332	<14354>	[7685]	4.119%	[95.347%]
6:	202	<15566>	[7887]	2.506%	[97.854%]
7:	77	<16105>	[7964]	0.955%	[98.809%]
8:	40	<16425>	[8004]	0.496%	[99.305%]
9:	29	<16686>	[8033]	0.360%	[99.665%]
10:	8	<16766>	[8041]	0.099%	[99.764%]
11:	8	<16854>	[8049]	0.099%	[99.864%]
12:	5	<16914>	[8054]	0.062%	[99.926%]
13:	3	<16953>	[8057]	0.037%	[99.963%]
14:	3	<16995>	[8060]	0.037%	[100.001%]

Total number of connectives	16995
Total number of sentences	8060
Average connectives per sentence	2.1086
Variance	2.9078
Standard deviation	1.7052
Coefficient of variation	80.8718%
Standard error	0.0190
Lower 95% confidence limit	2.0713
Upper 95% confidence limit	2.1458

Table 7.17 Connective-sentence distribution in the Arabic Corpus

Code of columns: Conn = Number of connectives at any one sentence
Sen = Number of sentences having numbers of connectives specified in column Conn.
Cum Conn = Cumulation of connective tokens (result of Conn x Sen)
Cum Sen = Cumulation of sentences
Rel Freq = Relative Frequency of Figures in Sen
Cum Freq = Cumulation of Figures in Rel Freq

2,608, which represents 32.35% of total sentences and 36.6% of sentences with connectives. They, thus, make up the largest portion of sentences containing connectives. Sentences that contain two connectives each come next in frequency. They make a total of 1983, representing 24.6% of total number of sentences and 27.8% of total sentences with connectives. Sentences having more than two connectives are 2536, which represents 31.4% of total sentences and 35.6% of sentences containing connectives. The last two percentages is, again, in sharp contrast with their counterpart in English (respectively 6.2% and 11.4%). It is interesting to note that there are three occurrences of a sentence containing 14 connectives.

The average connective per sentence in the Arabic corpus is 2.11. The confidence index shows that in 95% of occurrences the average connective per sentence ranges between 2.07 (the lower confidence limit) and 2.15 (the upper confidence limit). The standard deviation from this average is 1.7 and the coefficient of variation is calculated at 80.87%.

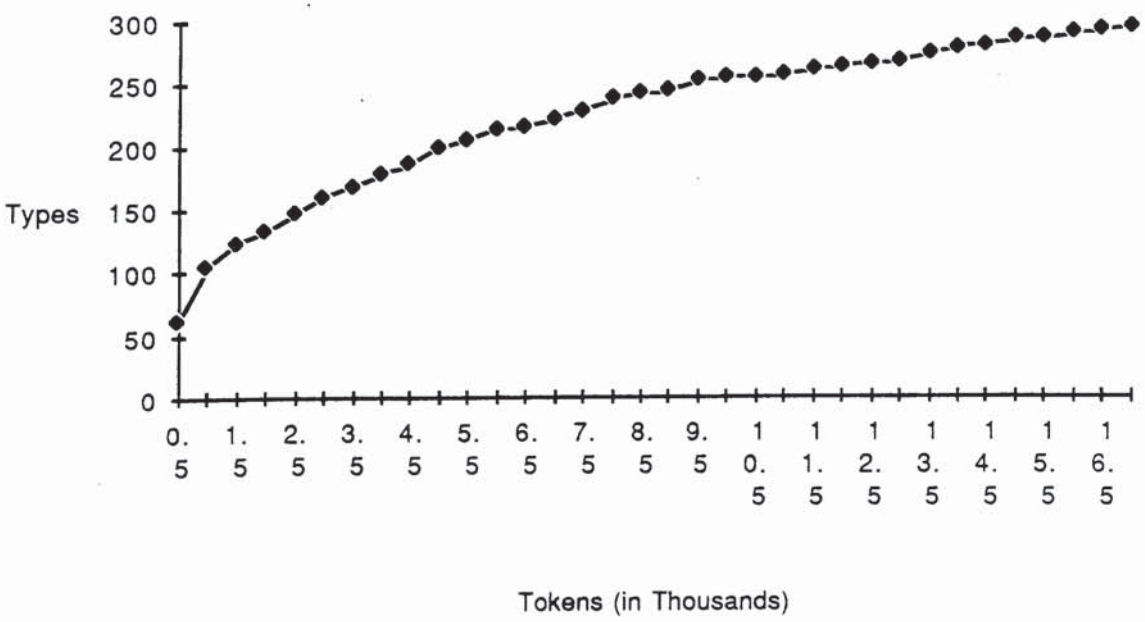
7.3.3.3 Type-token Measurements

Table (7.18) summarises results of the type-token measurements. Connectives in the Arabic corpus have a type-token ratio of 0.0174. The logarithmic TTR is 0.5839 and the double logarithmic TTR is 0.627. These indices, which represent richness and extensity of use of connectives, are substantially lower than their English counterpart. Figure (7.21) plots the types against the tokens in the corpus. The number of hapax legomena, as stated earlier, is 77, and that of the non-hapaxes is 220.

Concentration:	0.0130038
Consolidation factor:	0.741611
Density:	14.1168
Exclusivity:	0.00453074
Gravity:	220.714
Hapax probability:	0.258389
Non-hapax probability:	0.741611
Lexical efficiency:	0.484948
Lexical usage:	0.283635
Type-token ratio:	0.0175346
Predictability:	0.982465
Rhythmicality:	75.552
Stereotypicality:	76.552
Type occurrence rate:	57.0302
Variegation:	25.8389

Table 7.18 Indices based on connective type-token measurements in the Arabic corpus

Fig. 7.21 Distribution of Connective Types and Tokens in the Arabic Corpus



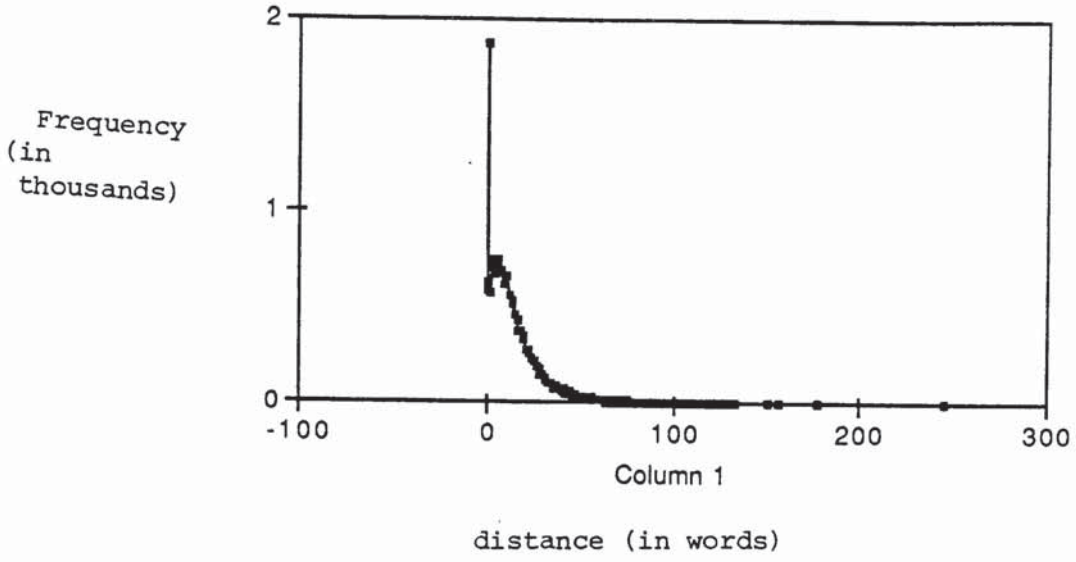
The concentration index is calculated at 0.013. This figure is lower than its English counterpart indicating a lower concentration of non-hapaxes in relation to the larger number of connective-tokens. The exclusivity index is 0.00453. The variegation index, which represents the percentage of connective diversification, is calculated at 25.84%. The hapax probability of occurrence is 0.258 while the non-hapax probability is 0.742. The gravity index, which indicates the rate of occurrence of hapaxes, is 220.7 a far bigger figure than the corresponding index in English owing to the higher number of connective tokens and lower number of hapaxes.

The type occurrence rate of connectives is 57, a larger figure than its counterpart in English. This difference is due to the larger number of tokens and fewer types. The density index is calculated at 14.1168. The stereotypicality index, which indicates the repeatedness of non-hapaxes, is computed at 76.55. The consolidation factor, an indication of the intensity of the repeatedness of non-hapaxes, is 0.742.

7.4.3.4 Entropy and Redundancy

The rank distribution is used for generating the entropy and redundancy indices. The full calculation of rank entropy is shown in Appendix (19). The maximum possible entropy, which represents the entropy of connectives when they are equally distributed, is 14.0528 bits. The entropy (H) of the connectives is 3.98587, a smaller index than its counterpart in the English corpus. The relative entropy (h), expressed as a percentage, is 28.3635%, far lower than its corresponding index in English. Redundancy is thus far greater, calculated as 71.6365%.

Fig. 7.22 Distribution of distance between occurrences of connectives in Arabic



7.4.3.6.2 Levin's Measure of Interval

The second measure applies Levin's formula for the characterisation of interval distribution of connectives. Results show that the L-compactness factor is 0.04632 and the Q-compactness factor is 0.00006. In general, these indices indicate, as in the English corpus, a tendency towards smoothness in the distribution of connectives throughout the text. The compactness factor, which is slightly lower than its counterpart in the English corpus, is indicative of a characteristic clustering of connectives.

7.4.3.7 Measure of Growth

Similar to the calculation of growth in the English corpus, two measures have been obtained: a "global", which monitors growth within intervals of running text, and "local", which computes growth

7.4.3.5 Repeat Rate Indices

Three indices of repeat rate have been computed to correspond to those of English connectives:

1. The general repeat rate (indicating the probability that two random selections from the entire Arabic corpus will give two connectives) is (44×10^{-4}), i.e. 44 in 10,000.

2. The connective system repeat rate (indicating the probability that two random selections from the entire Arabic corpus will yield the same connective) is (11×10^{-4}), i.e. 11 in 10,000.

3. The connective type repeat rate (indicating the probability that two random selections from the mass of connectives will yield the same connective type) is 0.25.

A full list of the repeat rate indices for each individual connective type is given in Appendices (22-23).

7.4.3.6 Measure of Interval

7.4.3.6.1 Gap Distribution

The first measure of interval calculates the distribution of gaps that separate the occurrences of connectives in the corpus. The results are given in Appendix (25) and represented in Figure (7.22). According to the distribution, there are 123 types of distance lengths. The most frequent one is 0 distance with a frequency of 1877, a large figure compared to its English counterpart. Distance lengths of one word has 627 occurrences. The longest distance length observed is 245 words, which occurs only once. The average length is 14 words.

of connective types within intervals of connective tokens. The ultimate aim is to compute the dependence of the number of connectives on text length and gain an insight in their manner of repetitiveness.

1. Global Growth

Results of these calculations are given in Table (7.19) and displayed in Figure (7.23). Connective tokens and types are computed in intervals of 5000 words of running text. Tuldava's formula is used for calculating growth rate, richness index, the expected growth within each corpus and, on the basis of these measurements, for extrapolation.

The first interval contains 369 connective tokens, which comprise 58 types. The expected figures are 362 tokens and 66 types. The second interval contains 358 more connective tokens, which include 31 new types. This raises tokens to 727 and types to 89.

Growth of tokens within the intervals varies from 421 (at interval 14), which is its highest level, to 220 (at interval 19), which is its lowest level. Note that the lowest level of growth of tokens is still higher than the highest level of growth in the English corpus.

The growth of types slows down after the first few intervals. There is total saturation, i.e. absence of any increase, of types at intervals 20, 33, 37, 44 and 47. The 50th percentile type occurs within interval 8 (40,000 words of running text); the 75th percentile type occurs within interval 22 (110,000 words); and the 90th percentile type occurs within interval 38 (190,000 words). These figures reflect a slower saturation of types in Arabic

---

Tuldava's index for connective tokens richness	2.009976
Growth rate of connective tokens:	-2.034383
Tuldava's index for connective type richness	0.356073
Growth rate of connective types:	2.015637

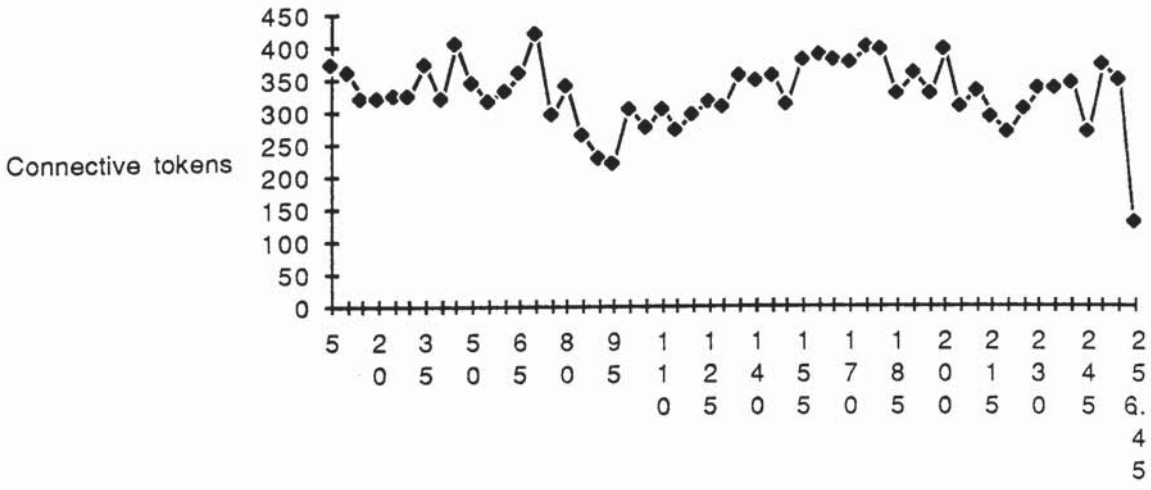
Actual Text Tokens	Actual Conn.Tokens	Expected Conn.Tokens	Actual Conn.Types	Expected Conn.Types
5000	369	362	58	66
10000	727	709	89	88
15000	1048	1051	108	103
20000	1368	1390	118	116
25000	1691	1727	132	126
30000	2015	2062	134	136
35000	2384	2396	145	144
40000	2703	2729	155	152
45000	3107	3060	161	159
50000	3452	3391	170	165
55000	3767	3721	175	172
60000	4100	4050	182	177
65000	4459	4379	187	183
70000	4880	4707	197	188
75000	5176	5035	204	193
80000	5516	5362	207	198
85000	5777	5688	214	203
90000	6005	6014	217	207
95000	6225	6340	219	211
100000	6528	6665	219	216
105000	6802	6990	220	220
110000	7106	7315	225	224
115000	7374	7639	229	227
120000	7668	7963	235	231
125000	7984	8287	238	235
130000	8291	8610	242	238
135000	8645	8933	246	242
140000	8994	9256	247	245
145000	9348	9579	253	248
150000	9660	9901	255	251
155000	10039	10223	257	254
160000	10424	10545	258	258
165000	10803	10867	258	261
170000	11176	11188	262	263
175000	11573	11509	263	266
180000	11968	11830	265	269
185000	12297	12151	265	272
190000	12652	12472	269	275
195000	12981	12792	271	277
200000	13376	13113	277	280
205000	13681	13433	280	283
210000	14011	13753	281	285

215000	14300	14073	284	288
220000	14566	14392	284	290
225000	14869	14712	286	293
230000	15204	15031	288	295
235000	15541	15350	288	297
240000	15884	15669	289	300
245000	16148	15988	291	302
250000	16519	16307	293	304
255000	16866	16625	296	307
256450	16995	16718	297	307

Extrapolated Text Tokens	Expected Connective Tokens	Expected Connective Types
500000	32113	393
750000	47752	455
1000000	63287	505
5000000	306514	894

Table 7.19 Calculation of growth of connectives in Arabic corpus

Fig. 7.23 Growth of connective tokens in the Arabic Corpus



compared to English, which is an important quantitative variation as will be discussed later (Ch. 9).

In extrapolating for connective tokens, Tuldava's richness index is calculated at 2.012733 and the connective growth rate is -2.036096. An extrapolated corpus of 500,000 is expected to include 32,138 connective tokens and 393 types. An extrapolated corpus of a million words is expected to contain 63,382 connective tokens and 505 types.

2. Local Growth

This measure monitors the growth of types within intervals of 500 connective tokens, regardless of the size of text that each interval may coincide with. Since there are 16,995 connective tokens, the corpus is divided into 34 intervals. The first comprises 63 types and gives a type-token ratio of 0.126, considerably lower than its counterpart in English. The expected number of types within the interval is 70. In the next interval, types grow by 43, raising the number of types to 106 and giving a ratio of 0.106. After this, the ratio decreases gradually, but at a lower rate than in English.

In extrapolating for connective types Tuldava's richness index is 0.010597 and the computed growth rate of types is 0.177492. An extrapolated number of 20,000 connective tokens (an increase of a little more than 3,000 on top of the existing figure) is expected to give 323 types (an expected growth of 28 types). Extending the extrapolated connective tokens to 30,000 is expected to yield 45 new types, raising the number of types to 368. An extrapolation of 50,000 connective tokens is expected to give 63 more types, bringing the total number of expected types to 431. An extrapolated number

Tuldava's richness index: 0.010597
Connective growth rate: 0.177492

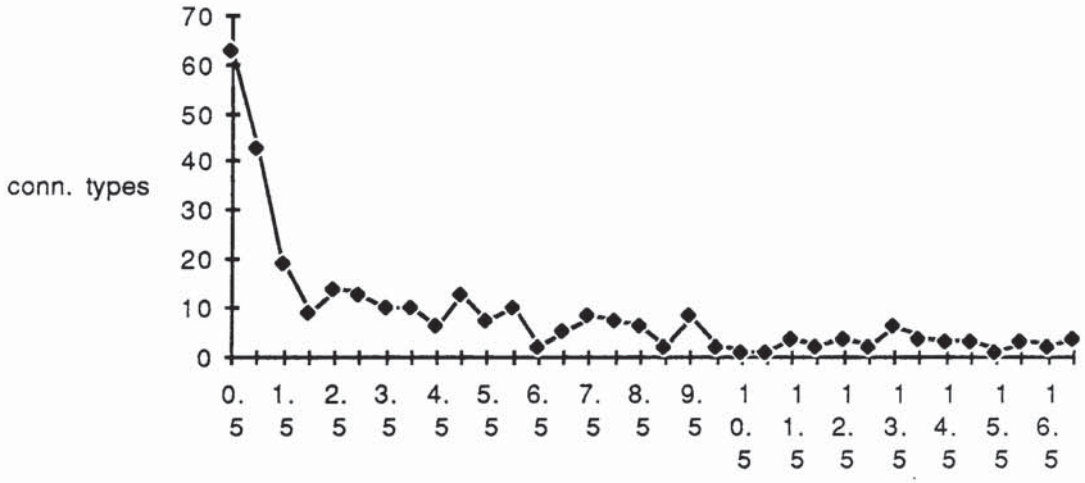
Actual Text Tokens	Actual Con.Toks	Actual Con.Types	Expected Con.Types
6874	500	63	70
14299	1000	106	98
22045	1500	125	118
29780	2000	134	133
36780	2500	148	147
43914	3000	161	158
50824	3500	171	169
58682	4000	181	178
65451	4500	187	187
72146	5000	200	195
79822	5500	207	202
89886	6000	217	209
99607	6500	219	216
108797	7000	224	222
116773	7500	232	228
125279	8000	239	233
133343	8500	245	239
140200	9000	247	244
147894	9500	255	249
154440	10000	257	253
161312	10500	258	258
167655	11000	259	262
174120	11500	263	266
180453	12000	265	270
188029	12500	269	274
195369	13000	271	278
202338	13500	277	282
210033	14000	281	285
219013	14500	284	289
227290	15000	287	292
234599	15500	288	296
242886	16000	291	299
249804	16500	293	302
256560	16995	297	305

Extrapolated Connective Tokens	Expected Connective Types
20000	323
30000	368
50000	431
100000	526

Table 7.20 Growth of connective types within connective tokens in the Arabic corpus

of 100,000 connective tokens is expected to produce 95 new types, raising the number of types to 526, which is distinctly higher than its English counterpart. The full results are given in Table (7.20) and displayed in Figure (7.24).

Fig. 7.24 Growth of Connective Types within Tokens in the Arabic Corpus



7.5 Conclusion

In this chapter, we have been concerned with elaborating the various components of our calculus of observations on connectives. The calculus is composed of two distinct profiles: a) a global statistical profile of each corpus; and b) a profile displaying the general quantitative characteristics of connectives in each corpus.

The description of these profiles is preceded by a brief survey of some statistical descriptions of connectives. It is our contention that these analyses are either based on simple word count procedures that give an insufficient, if not distorted, profile of connectivity, or provide a description of a too narrow conception of the phenomenon. In both cases, the analytical apparatus is related to a set of research aims that are different from ours.

To distinguish the statistical tools that we have employed in constructing the components of the calculus from those used in the surveyed studies, a short description of the typology of measurements is introduced in a concise and compact form. These measures are of five different types: a) measures related to frequency and rank distributions, b) measures based on the relations between types, tokens and hapax legomena, c) measures of entropy and redundancy, d) measures of repeat rate and interval distribution and e) measures of growth.

All calculations are computer-aided. Most of them are performed via a suite of SPITBOL programme, each designed to carry out a particular phase in the calculation. The rest of the calculations are obtained through various OCP runs. Each profile is then discussed with some detail, keeping the contrastive statements to the minimum (with the exception of the first component: the global profile of the two corpora).

The significance of the battery of measurements that we have used is that it has provided general profiles of various quantitative characteristics of connectives in the two corpora. Although we may still apply other measurements, such as establishing Yule's

characteristic K, Herdan's characteristic, measurement of "invisible" words, "invisibility" factor, Good's "D-1" and "D-2" characteristics, Brunet's richness index, Dolphin-Waring's various indices and Simpson's indices, we feel that the statistical observations we have gathered from our computational researches are sufficient to prove that the quantitative behaviour of connectives differs in both languages. The different statistical characteristics we have established underlie different text organisation in the two languages, a point that will be discussed in more detail later (Chapters 9 and 10). We will still, however, require to investigate the quantitative characteristics functional categories of the various connectives in each corpus. The calculus will therefore be extended to include a profile of each of the nine functional categories discussed in Chapter 6 above. This comprehensive task is left to the next chapter.

Footnotes to Chapter 7

(1) Some other statistical studies of connectives include Card (1972), Liskova (1983) and Lintermann-Rygh (1985).

(2) Nebeska's study is based on a project conducted by the members of the Department of Mathematical Linguistics under the leadership of Marie Tesitelova. See also Tesitelova (1980).

(3) The size of the journalistic text of the FDC is 137,201 lexical units. The relative frequency of conjunctions is 8.13; the relative frequency in Nebeska's corpus is 8.09.

(4) The fiction texts are taken from Dickens, Golding, D.H. Lawrence and Forster; texts of colloquial style are taken from Oscar Wilde, J. Osborne, Terence Rattigan, and J. Mortimer; the scientific texts are on astronomy, zoology, psychology and archaeology.

(5) When the semi-colon is included as a sentence delimiter, different figures for relative frequency of cohesive conjunctions are obtained. The genre most affected by this change is religion (see figures in Smith and Frawley 1983, p.357).

(6) Fiction and journalism use coordinating conjunctive cohesion more frequently than subordinating cohesion, but in religious and scientific texts the converse it true.

(7) Brainerd (1982) argues that the logarithmic type-token ratio (LTTR) is still sensitive to text length but drastically less so than the TTR. He therefore makes a number of suggestions to generalise the LTTR in a manner that would reduce some of its "systematic deviations".

(8) Herdan (1956, pp. 163 ff) discusses the method of calculating each of these entropies. He then maintains that

> "The reason why the entropy per symbol calculated in this way forms the appropriate measure of the amount of information is contained in the 'Fundamental Theorem' or the 'Coding Theorem'. This states that if a language has an entropy of H bits ... per phoneme or any other linguistic unit, then it is possible to approximate as closely as desired to it by a coding system which translates the original message into a binary code, with digits 0 and 1 say, using, on the average, H bits per linguistic unit of the original language. In accordance with this, H stands to the mean rank, M, of the original series in the relation ...:
>
> $$H \geqslant M.$$

(9) This term requires some explanation. If every symbol of the code (be it a phoneme, a letter, or other linguistic units) had the same probability of occurrence, there could be no guessing at missing parts of message since all signs are necessary or non-redundant. Conversely, a code is denoted as redundant if its signs

have unequal probabilities of occurrence and so admit guessing.

(10) There are a number of attempts to construct an appropriate formula which would express the dependence of vocabulary size upon text length. For instance, some formulae have been derived from theoretical considerations such as the hypothesis about the logarithmic distribution of words in text (cf. Carroll 1967). Others have been based on the Zipf law (for instance Kalinin 1964 and Orlov 1976 reported in Tuldava 1977). Earlier formulae have been suggested by a number of scholars such as Guiraud (1959) and Herdan (1964). Guiraud uses the hypothesis about the connection between the size of vocabulary and the square root of the number of word occurrences. Herdan uses the formula of the rank function of type $L = aN^b$ (where a and b are constants).

(11) According to Tuldava (1977 p.32) the maximum value is reached at

$$N_{max} = 10^{1/\sqrt[b]{a(b+1)}}$$

(12) See Knowles (1987) for a brief discussion and application.

CHAPTER EIGHT

The Quantification of Connectives: Calculus of Functional Categories

8.0 Perspective

The measurements that were adopted in the previous chapter were applied on connectives regardless of their semantic functioning in the text. The resulting calculus is therefore applicable only in general terms to text organisation. However, connectives are one of several resources that sustain textual cohesion (see arguments in Chapter 4) and cohesion is fundamentally a semantic concept (see Chapter 3). The basic role of connectives is to help create text by relating and organising text sequences in a rhetorically meaningful and effective way. The operationality of connectives is dependent on the semantic/rhetorical expression of relationships within text and these have been categorised as functions. Since we have already displayed that the behaviour of connectives vary statistically in both English and Arabic and that the variation is reflected nearly in every quantitative characteristic investigated, it follows that the quantitative nature of the functional categories are responsive to these variations. A requirement is, therefore, set up for investigating these variations and assessing their extent in both languages.

The aim of this chapter is then to provide a calculus of observations on the characteristics of each functional category of connectives. This aim is achieved by a battery of measurements that are intended to calculate the following:

a. Frequency distribution of the connectives in the category including a discussion of the rank distribution.

b. Frequency distribution of each subcategory.

c. Indices of repetitiveness including category token and type occurrence rate, repeat rate and gap distribution.

d. Growth rate and extrapolation.

To avoid prolonging the chapter unnecessarily, we shall provide only brief profiles for subcategories. Detailed frequency lists, offering absolute, relative and cumulative frequencies, are provided for reference in Volume (4).

It cannot be overemphasised that computation of connectives based on differentiation of functional meaning will drastically affect the statistical distribution of connectives. A connective such as "and" with a total frequency of 1720 occurrences has several functional meanings in the text: it is basically additive, but can also be temporal, causal and adversative. The distribution of the occurrences of "and" over these categories reduces its frequency at any one count, while at the same time creating four different connective types from one connective token.

This statistical profile for functionality complements the general statistical profile of connectives given earlier. Both proffer a comprehensive description that assists us in assessing and understanding how connectives operate in each language. And this is the ultimate aim in this project.

8.1 A Statistical Preview of Functionality

Before we give the details of the statistical characteristics of each category, we would like to give a brief account of the

statistical composition of each category in terms of types and tokens. This will be offered first for English and then for Arabic.

8.1.1 Profile of Functional Categories in English

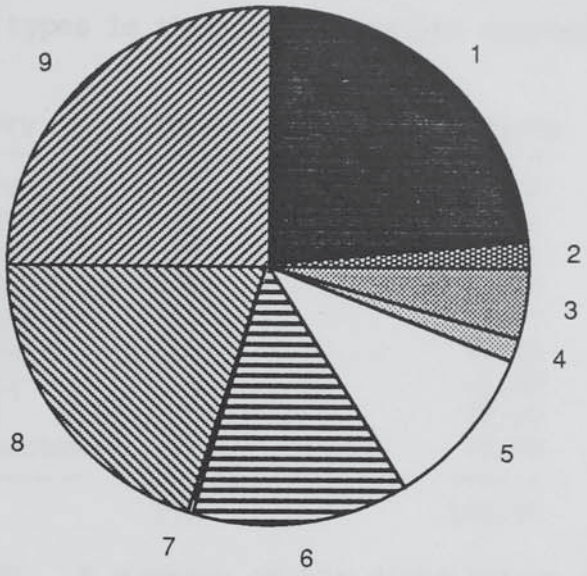
Tables (8.1) and (8.2) display the type-token counts of each of the nine categories of connectives in English. Table (8.1) gives the absolute, relative and cumulative frequencies of tokens within each category. According to this table, the largest category of connectives is adversative. It comprises 2,404 tokens representing more than 25% of the total connective mass. This category is followed closely by additives, which consist of 2,260 tokens representing 23.55% of total connectives. The smallest category is spatial connectives, which consists of 50 tokens representing only 0.5% of total connectives. Figure (8. 1) displays the share that each category has of connective mass.

Category	Absolute Frequency	Relative Frequency	Cumulative Frequency
Additive	2260	23.55	23.55
Alternative	143	1.49	25.04
Comparative	441	4.60	29.64
Reformulatory	115	1.20	30.84
Orientative	1011	10.54	41.38
Temporal	1284	13.38	54.76
Spatial	50	0.52	55.28
Causal	1888	19.67	74.95
Adversative	2404	25.05	100.00
Total	9596	100.00	

Table 8.1 A summary account of the size of connective categories in terms of tokens

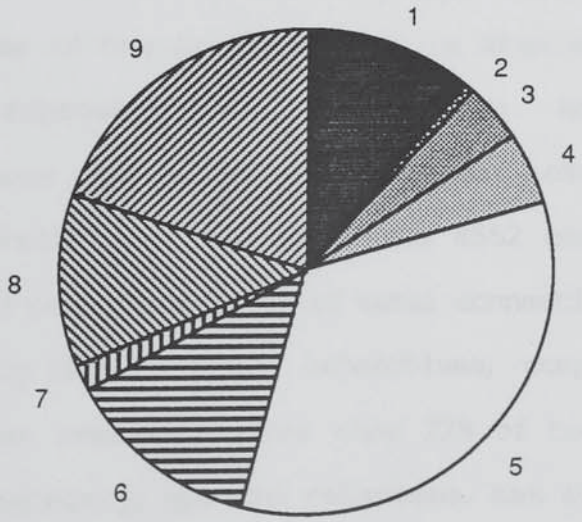
Table (8.2) summarises distribution of types within categories. As mentioned earlier (Ch. 5) types computed here represent connective cores.[1] The total number of categorised connective cores is 375, giving an increase of 64 types over general untagged

Fig. 8.1 Distribution of Categories of Connectives in the English Corpus



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

Fig. 8.2 Distribution of Connective Types within the Categories of Connectives in the English Corpus



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

connective cores. The largest number of types occur within the category of orientative connectives. It includes 127 types that represent approximately 34% of total types. The smallest category in terms of types is that of alternative connectives. It comprises

Category	Types (cores)	% Types
Additive	42	11.20
Alternative	2	0.53
Comparative	15	4.00
Reformulatory	18	4.80
Orientative	127	33.87
Temporal	48	12.80
Spatial	6	1.60
Causal	45	12.00
Adversative	72	19.20
Total	375	100.00

Table (8.2) A summary of the distribution of types within the functional categories of connectives in the English corpus

two types that stand for 0.53% of total connective types. Figure (8.2) displays the distribution of gaps within the categories.

8.1.2 Profile of Functional Categories in Arabic

A statistical summary of the type-token count of categorised connectives in the Arabic corpus is displayed in Table (8.3) and (8.4) and represented in Figures (8.3-4). According to Table (8.3) the category that achieves the largest token coverage is that of additive relations. It comprises 8552 connective tokens that correspond to more than 50% of total connectives. The next largest category is that of causal connectives, comprising 3778 connective tokens that represent more than 22% of total connectives. The smallest category, spatial relations, has 40 tokens which make up 0.24% of total connectives.

Category	Absolute Frequency	Relative Frequency	Cumulative Frequency
Additive	8552	50.32	50.32
Alternative	196	1.15	51.47
Comparative	266	1.57	53.04
Reformulatory	239	1.41	54.45
Orientative	738	4.34	58.79
Temporal	1494	8.79	67.58
Spatial	40	0.24	67.82
Causal	3778	22.23	90.05
Adversative	1692	9.95	100.00
Total	16995	100.00	

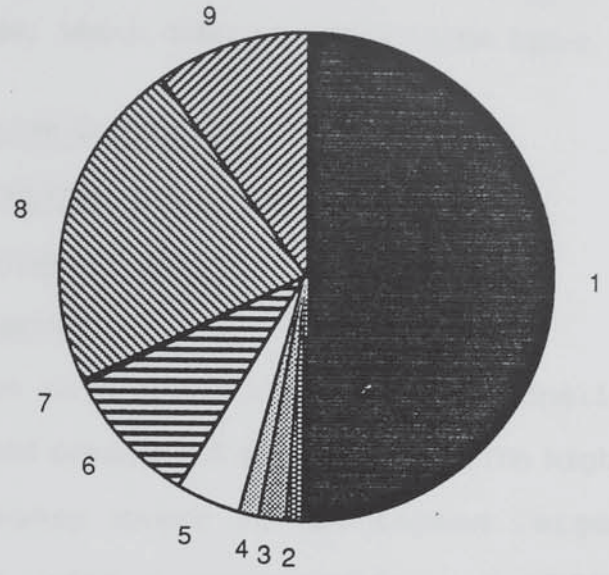
Table (8.3) A summary of the size of connective categories in terms of tokens in the Arabic corpus

Table (8.4) gives results of computing connective types within each category. The total number of categorised type cores is 382. This figure represents an increase of 85 types (equal to 29%) over untagged connectives. The size of the increase is bigger than its counterpart in English (which is 21% of total untagged types) and this difference is related to a bigger number of multi-functional connectives in Arabic, a point that will be picked up later (Ch.9).

Category	Types (cores)	% Types
Additive	53	13.87
Alternative	4	1.05
Comparative	12	3.14
Reformulatory	27	7.07
Orientative	92	24.08
Temporal	62	16.23
Spatial	3	0.79
Causal	76	19.90
Adversatives	53	13.87
Total	382	100.00

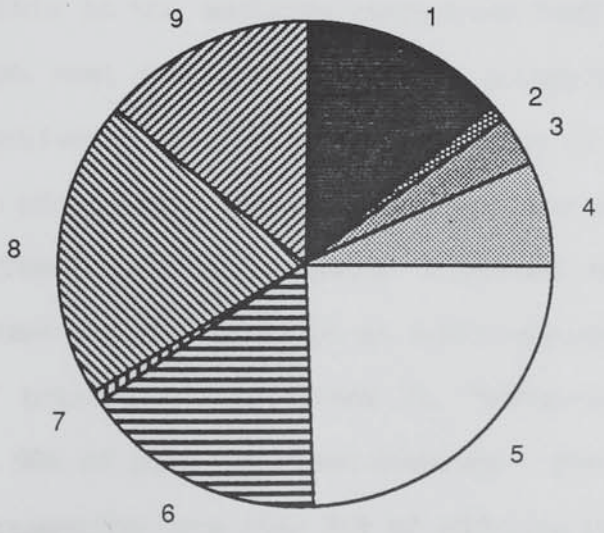
Table (8.4) A summary of the distribution of types within the functional categories of connectives in the Arabic corpus

Fig. 8.3 Distribution of Categories of Connectives in the Arabic Corpus



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

Fig. 8.4 Distribution of Connective Types within Categories of Connectives in the Arabic Corpus



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

According to the table, the largest category in terms of types is that of orientative connectives, which comprises 24% of total categorised types. The smallest category is that of spatial connectives, which comprises only three types.

8.2 Additive Connectives

8.2.1 Additive Connectives in English

8.2.1.1 Global Profile

1. Frequency

Additive connectives observed in the English corpus comprise a set of 2,260 tokens and 42 type cores. The high frequency figure of this category makes it the second largest among connective categories, with a token coverage of 23.55%. Alphabetical and frequency lists are given in App. 28 and 29. The distribution is plotted in the figures in App. 30-31. The rank distribution in Table (8.5) shows that the top rank is occupied with one connective with a frequency occurrence of 1,359, representing more than 60% of additive token coverage and more than 14% of total connective tokens. This is the additive connective "and".

The next most frequent connective, occupying the second rank, is the connective "also". It has a frequency of 290 which represents about 12% of additive connective tokens and 3% of total connective tokens. These two connectives, together with "even" (rank 3), "again" (rank 4), "which" - in an additive-commentative sense (rank 5), "too" (rank 6), "nor" (rank 7), "after-all" (rank 8) achieve more than 90% of additive token coverage. The rest of the ranks (34 types representing more than 80% of additive types) achieve only 10% of additive occurrences.

Rank	Tokens	Types	Rank	Tokens	Types
1:	1359	1	12:	14	1
2:	290	1	13:	13	1
3:	142	1	14:	10	2
4:	76	1	15:	9	2
5:	63	1	16:	7	2
6:	46	1	17:	6	2
7:	41	1	18:	5	1
8:	34	1	19:	4	1
9:	31	1	20:	3	5
10:	23	1	21:	2	5
11:	21	1	22:	1	9

Number of "Additive" connective tokens: 2260
Number of "Additive" connective types : 42

Table 8.5 Rank distribution of "additive" connectives in the English corpus

The lowest rank is occupied by 9 hapaxes. These are "added to that", "to begin with", "beyond that", "initially", "last", "more surprisingly" (used in an additive sense), "third", "on top of that", and "ultimately". (The underlined words represent the core elements in multi-word connectives.) The combined frequency of this rank makes only 0.4% of additive tokens but more than 21% of additive connective types.

2. Repetitiveness

Computation of repetitiveness of this category in the corpus has yielded the following indices.

a. The additive occurrence rate is 4.25. This index means that an additive is encountered within 4.25 successive occurrences of connectives with or without gaps separating them.

b. Type occurrence rate within the additive category is

approximately 54. This means that one type occurs in every 54 successive additive connectives.

c. General additive repeat rate, referring to the probability that two successive connectives - with or without a gap separating them - are any additive connectives, is (55×10^{-3}), i.e. 55 in a thousand.

d. The additive system repeat rate, referring to the probability that two successive connectives - with or without a gap separating them - are the same additive connective, is (21×10^{-3}).

e. The gap distribution indicates that he average gap (in terms of words) that separates two sites occupied by additives is 112.4 words. A gap of 0 (the two sites follow immediately one another) is the most frequent in the distribution with 35 occurrences. The largest gap observed in the corpus is 1,397 words.

3. Growth

Two calculations of growth have been made and the results are given in Tables (8.6-7) and plotted in the figures in App. (32-33). The first calculation is global: it monitors the growth of additives within the entire corpus. The second calculation of growth is local: it monitors the growth of additive connectives within connective tokens.

a. Global Growth: Global growth is investigated in text intervals of equal size, each 5,000 words long. This means that the corpus of 256,560 words is divided into 51 intervals of equal length plus a short interval at the end comprising 1,560 words. Growth in this short interval will not be overlooked in the discussion and comparison.

The first 5,000 of running text contains 43 additive connective tokens that contain 9 types. The next 5,000 words includes 50 more additive connective tokens but only 3 extra types. According to the results of calculating Tuldava's formula, the growth rate of additive connective tokens in the corpus is -4.060188 while growth rate of types is -8.000743 (the negative indices are generated by calculating the log-log numbers of small values). Extrapolation of a million word corpus is expected to include 9,309 additive tokens and 74 types.

b. Local Growth: Here growth is examined in text intervals each comprising 500 connective tokens, regardless of the size of running text that this interval may coincide with. Since the English corpus includes 9,596 connective tokens, it is divided into 19 intervals plus a short one that contains only 96 tokens. For convenience and accuracy of comparison, growth in this last interval will be overlooked.

Tuldava's (token) richness index:	4.005561
Growth rate of "additive" tokens:	-4.060188
Tuldava's (type) richness index:	0.902086
Growth rate of "additive" types:	-8.000743

Actual Text Tokens	Actual Add.Tokens	Expected Add.Tokens	Actual Add.Types	Expected Add.Types
5000	43	47	9	10
10000	93	94	12	13
15000	143	142	17	15
20000	185	189	17	17
25000	232	236	20	18
30000	281	283	21	20
35000	345	330	23	21
40000	385	377	25	22
45000	413	423	25	23
50000	464	470	26	24

55000	513	517	26	25
60000	567	564	26	25
65000	631	611	27	26
70000	685	658	27	27
75000	730	705	28	28
80000	774	751	29	28
85000	828	798	30	29
90000	881	845	32	30
95000	940	892	32	30
100000	984	938	32	31
105000	1036	985	32	31
110000	1079	1032	33	32
115000	1128	1079	34	32
120000	1161	1125	34	33
125000	1207	1172	35	33
130000	1254	1219	35	34
135000	1310	1266	35	34
140000	1352	1312	36	35
145000	1401	1359	36	35
150000	1452	1406	36	36
155000	1478	1452	36	36
160000	1524	1499	36	37
165000	1558	1546	36	37
170000	1600	1592	36	38
175000	1636	1639	37	38
180000	1680	1686	39	38
185000	1713	1732	40	39
190000	1759	1779	40	39
195000	1816	1826	40	40
200000	1853	1872	40	40
205000	1879	1919	40	40
210000	1907	1966	40	41
215000	1938	2012	41	41
220000	1984	2059	41	41
225000	2028	2105	41	42
230000	2062	2152	41	42
235000	2100	2199	42	42
240000	2135	2245	42	43
245000	2181	2292	42	43
250000	2214	2338	42	43
255000	2245	2385	42	44
256560	2260	2400	42	44

Extrapolated Text Tokens	Expected Add. Tokens	Expected Add. Types
500000	4666	57
750000	6989	66
1000000	9309	74

Table 8.6 Global Growth of Additive Connectives in the English Corpus

Tuldava's richness index:	1.011012
"Additive" Connective growth rate:	-1.024956

Actual Connective Tokens	Actual "Additive" Connective Tokens	Expected "Additive" Connective Tokens
500	129	128
1000	245	252
1500	377	375
2000	474	497
2500	626	618
3000	746	739
3500	877	860
4000	1003	980
4500	1130	1100
5000	1246	1220
5500	1381	1340
6000	1479	1460
6500	1584	1579
7000	1692	1698
7500	1821	1817
8000	1919	1936
8500	2037	2055
9000	2128	2174
9500	2237	2293
9596	2260	2315

Extrapolated Connective Tokens	Expected "Additive" Connective Tokens
10000	2411
15000	3592
20000	4767
30000	7106
50000	11751
100000	23267

Table 8.7 Local growth of additive connectives in the English corpus

The first interval (of 500 connective tokens) include 129 additive connectives, i.e. more than 25% of the connectives of that interval. This percentage fluctuates slightly as connectives grow in the corpus. Tuldava's richness index is calculated at 1.011012 and the growth rate of additive connectives is -1.024956 (the negative sign is a result of using logarithms of figures of low value). Extrapolation of 10,000 connective tokens is expected to have 2,411 additive connectives and an extrapolated number of 100,000 connective tokens is estimated to comprise 23,267 additives.

8.2.1.2 Categories of Additive Connectives in English

The distribution of the categories of additive connectives is displayed in Table (8.8) and Figure (8.5) (see also lists of frequency and rank distribution in App. 34A-E). The total number of types (in terms of cores) in the subcategories is 50, representing an increase of 8 types to overall additive types (which are 42). This increase (in this as well as other categories) is due to two reasons:

a. Certain additive connectives occur as types in more than one subcategory. For instance, the connective "and" stands for 4 types: as an appending, continuative, enumerative and commentative.

b. Certain connectives may occur as two types of core: as a core of a simple connective and as a core of a compound connective, and hence represents at this level of categorisation two distinct connectives. For instance "again" has been treated as a one-word connective and as a core in the compound connective "once again". (For a list of full-form types in the categories see App. 28).

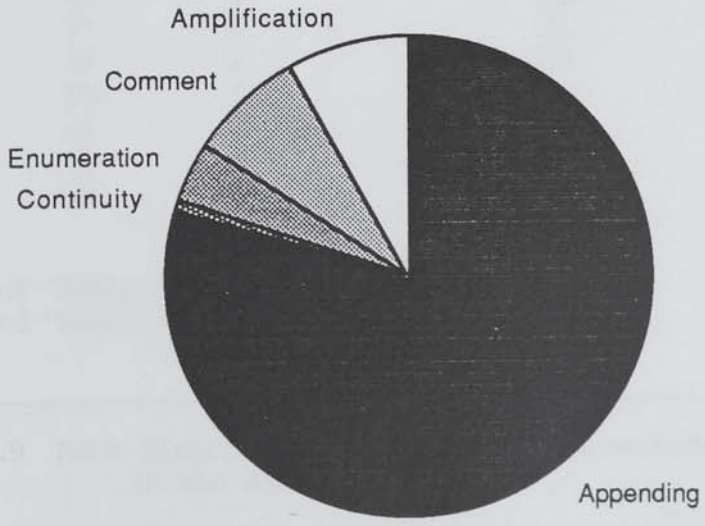
Inspection of Table (8.8) shows that the main additive category is that of "appending". It comprises about 80% of tokens and 42% of types. The next category in terms of size of token coverage is that

Category	% Tokens	% Tokens	Types (cores)	% Types
Appending	1805	79.87	21	42
Continuity	16	0.71	1	2
Enumeration	88	3.89	18	36
Comment	173	7.65	3	6
Amplification	178	7.88	7	14
	2260	100.00	50	100.00

Table (8.8) Distribution of Additive Subcategories in the English Corpus

of "amplification" followed closely by "comment". In terms of types the next big category is that of enumerative connectives: 18 types representing 36% of additive types. The smallest category both in

Fig. 8.5 Distribution of Categories of Additive Connectives in the English Corpus



terms of types and tokens is that of continuatives: 16 observed occurrences for the connective "and".

8.2.2 Additive Connectives in Arabic

8.2.2.1 Global Profile

1. Frequency

The additive category of connectives is by far the largest in the corpus in terms of frequency of occurrence. Additive connectives (with 8,552 tokens and 53 types) represent more than 50% of total connectives. The rank distribution (Table 8.9) (see also the alphabetical and frequency lists in Appendices 35 and 36) contains 25 ranks. The top rank is occupied by the connective "wa", which

Rank	Tokens	Types	Rank	Tokens	Types
1:	7021	1	14:	16	1
2:	619	1	15:	11	1
3:	163	1	16:	10	2
4:	161	1	17:	9	1
5:	132	1	18:	8	1
6:	83	1	19:	7	1
7:	43	1	20:	6	2
8:	37	1	21:	5	3
9:	36	1	22:	4	3
10:	29	1	23:	3	7
11:	21	1	24:	2	4
12:	19	1	25:	1	13
13:	18	2			

Number of "Additive" connective tokens: 8552
Number of "Additive" connective types : 53

Table 8.9 Rank distribution of additive connectives in the Arabic corpus

has a frequency of 7,021 representing 82% of total additive occurrence and more than 41% of total connectives in the corpus. The second top rank is filled by the connective "fa" with 619 occurrences representing slightly more that 7% of total additive connectives. Thus these two high frequency connectives correspond to more than 89% of additive occurrences. The frequency of the rest of the additive connectives (51 types representing 96% of additive types is equivalent to less than 11% of total additive occurrences. The lowest rank is occupied by 13 hapaxes that correspond to about 24.5% of additive types but only 0.15% of total additive tokens.

2. Repetitiveness

Computation of repetitiveness of connectives of this category has yielded the following indices:

a. The additive occurrence rate is 2, indicating that in every 2 connective tokens one is an additive. This is a high rate which signals an important characteristic of textual organisation in Arabic.

b. Type occurrence rate is 161.4. This means that we encounter a new additive type in every 161.4 occurrence of additive tokens. This low rate is due to the enormity of the number of the Arabic additive tokens in relation to types.

c. General additive repeat rate, referring to the probability that two successive connectives - with or without a gap separating them - are any additives is 0.25.

d. Additive system repeat rate, referring to the probability that two successive connectives - with or without a gap separating them - are the same additive connective, is 0.172.

e. The gap distribution reveals that the average gap that separates two sites occupied by additive connectives is about 29 words. A gap of 0 words (two additive connectives follow each other immediately) occurs 152 times. The most frequent gap has an occurrence number of 254 and comprises 15 words. The next most frequent gap comprises 10 words and has a frequency of 230. The largest gap has one occurrence and comprises 305 words.

3. Growth

Two calculations of growth have been made and the results are displayed in Tables (8.10-11). Growth is plotted in App. (39) and (40). The first calculation measures global growth of additives; the second computes growth of additives in relation to overall connective tokens.

a. Global Growth: This measure monitors the growth of additive types and tokens within text intervals of 5,000 words of running text each. The corpus is therefore divided into 52 intervals as displayed in Table (8.10). Tuldava's formula is also applied to compute the growth rate and to extrapolate growth of types and tokens within larger corpora.

The first 5,000 words comprises 185 additive tokens which contain 7 different additive types. The expected number of tokens is 178, slightly lower, but the expected number of types is higher, 14. The interval comprises 169 additive tokens that include 15 new types. This is the biggest number of additive types within any of the 52 text intervals. Growth of tokens remains fairly consistent. The largest number of additive tokens observed within any interval is 205 at interval 7. In contrast, growth of types slows down

---

Tuldava's (token) richness index:	2.096313
Growth rate of "additive" tokens:	-3.013509
Tuldava's richness index:	0.768856
Growth rate of "additive" types:	-14.001305

Actual Text Tokens	Actual Add.Tokens	Expected Add.Tokens	Actual Add.Types	Expected Add.Types
5000	185	178	7	14
10000	354	351	22	18
15000	522	523	26	21
20000	664	694	27	24
25000	834	864	30	26
30000	1004	1034	32	27
35000	1209	1203	34	29
40000	1372	1371	36	30
45000	1574	1540	36	32
50000	1744	1708	36	33
55000	1889	1875	36	34
60000	2056	2043	36	35
65000	2229	2210	36	36
70000	2424	2377	41	37
75000	2593	2544	42	38
80000	2786	2711	43	39
85000	2917	2878	44	40
90000	3037	3044	46	41
95000	3168	3211	46	42
100000	3338	3377	46	43
105000	3497	3543	46	43
110000	3673	3709	47	44
115000	3806	3875	47	45
120000	3957	4040	48	46
125000	4126	4206	49	46
130000	4281	4372	49	47
135000	4446	4537	50	48
140000	4646	4702	50	48
145000	4817	4868	50	49
150000	4986	5033	50	49
155000	5187	5198	51	50
160000	5369	5363	51	51
165000	5546	5528	51	51
170000	5714	5693	51	52
175000	5891	5858	51	52
180000	6074	6022	52	53
185000	6253	6187	52	53
190000	6435	6352	52	54
195000	6606	6516	52	54
200000	6777	6681	52	55
205000	6903	6845	53	55
210000	7071	7010	53	56

215000	7206	7174	53	56
220000	7345	7338	53	57
225000	7520	7503	53	57
230000	7681	7667	53	58
235000	7843	7831	53	58
240000	8025	7995	53	59
245000	8149	8159	53	59
250000	8339	8323	53	60
255000	8500	8487	53	60
256450	8552	8535	53	60

Extrapolated Text Tokens	Expected Add. Tokens	Expected Add. Types
500000	16479	77
750000	24578	90
1000000	32641	43

Table 8.10 Global Growth of Additive Connectives in the Arabic Corpus

considerably after the first two intervals until types are saturated at interval 41, which represents 205,000 word level. Extrapolation of a corpus of 500,000 words of running text is expected to yield 16,479 additive tokens and 77 types. Extrapolation to a million word corpus is expected to include 32,641 additive tokens and types.

b. Local Growth: This measure computes the growth of additive connectives within intervals of connective tokens. The first 500 connective tokens contain 239 additive tokens. The expected number, arrived at by using Tuldava's formula, is 244 tokens and types. Growth rate of additive tokens within connectives is -0.819363 and richness index is 0.941824. Extrapolation of 20,000 connective tokens is expected to contain 10,260 additives and an extrapolated number of 100,000 connectives is estimated to yield 52,012.

Tuldava's richness index:	0.941824
"Additive" Connective growth rate:	-0.819363

Actual Connective Tokens	Actual "Additive" Connective Tokens	Expected "Additive" Connective Tokens
500	239	244
1000	504	493
1500	726	745
2000	998	997
2500	1264	1250
3000	1520	1504
3500	1765	1758
4000	2003	2013
4500	2238	2268
5000	2482	2523
5500	2775	2778
6000	3034	3034
6500	3324	3290
7000	3615	3546
7500	3868	3803
8000	4134	4059
8500	4377	4316
9000	4647	4573
9500	4904	4830
10000	5168	5087
10500	5414	5344
11000	5628	5601
11500	5854	5859
12000	6097	6116
12500	6354	6374
13000	6617	6632
13500	6831	6889
14000	7066	7147
14500	7303	7405
15000	7596	7664
15500	7823	7922
16000	8076	8180
16500	8329	8438
16995	8552	8694

Extrapolated Connective Tokens	Expected "Additive" Connective Tokens
20000	10248
30000	15434
50000	25848
100000	52012

Table 8.11 Local growth of additive connectives in the Arabic corpus

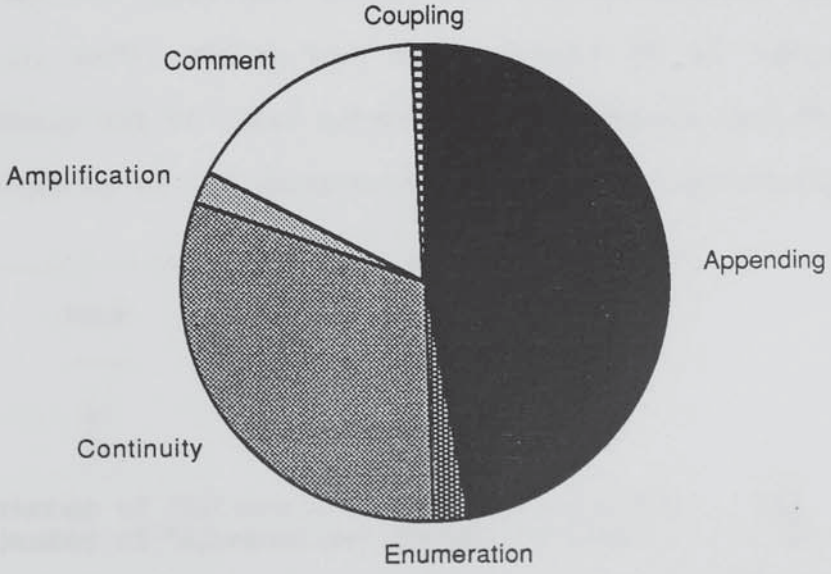
8.2.2.2 Categories of Additive Connectives in Arabic

Results of the calculation of additive categories are displayed in Table (8.12). A more detailed frequency lists are given in App. (41A-E). The share that each additive category has in token mass is

Category	Tokens	% Tokens	Types	% Types
Appending	4058	47.45	21	32.81
Continuity	2676	31.29	7	10.94
Enumeration	180	2.10	20	31.25
Comment	1380	16.14	5	7.81
Amplification	204	2.39	8	12.50
Coupling	54	0.63	3	4.69
Total	8552	100.00	64	100.00

Table (8.12) Distribution of Additive Subcategories in the Arabic corpus

Fig. 8.6 Distribution of the Categories of Additive Connectives in the Arabic Corpus



displayed in Figure (8.6). The total number of types in the subcategories is 64, representing an increase of 11 types over general additive types, i.e. an increase of 21%.

The largest additive category is that of "appending". It comprises 22 types that have a total frequency of 4,058 representing about 47.5% of additives. The next biggest additive category in terms of token coverage is that of continuatives. It has a frequency of 2,627 corresponding to 31% to additives. In terms of types, the next biggest additive category is that of enumerative, having 20 types. The smallest category is that of coupling; it comprises 3 types that has a frequency of 54 tokens.

8.3 Alternative

8.3.1 Alternative Connectives in English

This is a relatively small category of connectives with only two types that have a total frequency of 143. The type-token ratio is 0.014. The two types are "either" (in its correlative construction "either ... or"), which has a frequency of 17 representing approximately 12% of total alternative connectives, and "or", which has a frequency of 126 representing 88% of total alternative tokens.

Rank	Tokens	Types
1:	126	1
2:	17	1

Number of "Alternative" connective tokens: 143
Number of "Alternative" connective types : 2

Table 8.13 Rank distribution of "Alternative" connectives in the English Corpus

The rank distribution is give in Table (8.13) and the frequency lists in App. (42). In addition, App. (43) and (44) plot the distribution of alternative connectives against text tokens and overall connective tokens.

2. Repetitiveness

Computation of repetitiveness of this category in the corpus has yielded the following indices:

a. The occurrence rate of alternative connectives is 67. This indicates that an alternative connective is encountered in each set of 67 connectives.

b. Type occurrence rate is 71.5.

c. General repeat rate indicating the probability that two successive connectives - with or without a gap - are alternatives is (2×10^{-4}), i.e. 2 in 10,000.

d. Alternative system repeat rate is approximately (18×10^{-5}). This refers to the probability that two successive connectives are the same alternative connective.

e. Gap distribution shows that the average gap that separates two occurrences of alternative connectives is 1,780 words. The smallest gap comprises one word and has a frequency of 2. The largest gap consists of 7,333 words.

3. Growth

Two calculations of growth are performed: global and local. These are given in Tables (8.14-15) and displayed in the diagrams in Appendices (45) and (46).

Tuldava's (token) richness index: 10.002679
Growth rate of "alternative" tokens: -9.000807
Tuldava's (type) richness index: 1.018265
Growth rate of "alternative" types: -12.000794

Actual Text Tokens	Actual Alt.Tokens	Expected Alt.Tokens	Actual Alt.Types	Expected Alt.Types
5000	1	2	1	1
10000	4	4	1	1
15000	10	7	1	1
20000	12	9	2	1
25000	14	12	2	1
30000	18	15	2	1
35000	20	18	2	1
40000	21	21	2	1
45000	24	24	2	1
50000	29	27	2	1
55000	32	30	2	1
60000	35	33	2	1
65000	38	36	2	1
70000	45	39	2	1
75000	48	42	2	1
80000	52	45	2	1
85000	54	48	2	1
90000	55	51	2	1
95000	58	54	2	1
100000	63	57	2	1
105000	66	60	2	1
110000	70	63	2	1
115000	74	66	2	1
120000	78	70	2	1
125000	79	73	2	1
130000	81	76	2	1
135000	82	79	2	2
140000	85	82	2	2
145000	88	86	2	2
150000	89	89	2	2
155000	90	92	2	2
160000	95	95	2	2
165000	96	98	2	2
170000	98	102	2	2
175000	100	105	2	2
180000	102	108	2	2
185000	102	112	2	2
190000	104	115	2	2
195000	108	118	2	2
200000	112	121	2	2
205000	113	125	2	2
210000	117	128	2	2
215000	121	131	2	2

220000	130	135	2	2
225000	131	138	2	2
230000	133	141	2	2
235000	134	144	2	2
240000	138	148	2	2
245000	140	151	2	2
250000	141	154	2	2
255000	142	158	2	2
256560	143	159	2	2

Extrapolated Text Tokens	Expected Alt. Tokens	Expected Alt. Types
500000	326	2
750000	504	2
1000000	686	2

Table 8.14 Global growth of alternative connectives in the English corpus

a. Global Growth: This measure traces the growth of alternative connectives within intervals of 5000 words. The first interval includes only one occurrence, which is the same as the expected number, arrived at by using Tuldava's measure.

The next interval contains three more occurrences. The largest number of alternatives appearing at any interval is 9, which appears in interval 44. Extrapolation of a corpus of 500,000 words is expected to contain 326 alternative tokens but no increase in types. Extrapolation of a million word corpus is expected to include 686 alternative connective tokens, but the same two alternative types.

b. Local Growth: This measure traces the growth of alternative connectives within intervals of 500 successive connectives each. The first interval contains only 6 alternative connectives. The next 500 connectives has 9 alternatives, bringing the total to 15 in

the first 1,000 connective tokens. An extrapolated number of 10,000 connectives gives 156 tokens. An extrapolation of 100,000 connective tokens is expected to render 1,691 alternative tokens.

Tuldava's richness index:	5.002489
"Alternative" Connective growth rate:	-4.006311

Actual Connective Tokens	Actual "Alternative" Connective Tokens	Expected "Alternative" Connective Tokens
500	6	6
1000	15	14
1500	21	21
2000	29	29
2500	38	36
3000	48	44
3500	55	52
4000	64	60
4500	74	67
5000	80	75
5500	87	83
6000	90	91
6500	96	99
7000	102	107
7500	108	115
8000	121	123
8500	131	131
9000	137	139
9500	142	147
9596	143	149

Extrapolated Connective Tokens	Expected "Alternative" Connective Tokens
10000	156
15000	237
20000	320
30000	486
50000	825
100000	1687

Table 8.15 Local growth of alternative connectives in the English corpus

8.3.2 Alternative Connectives in Arabic

1. Frequency

Connectives of alternative relations constitute a small category in the Arabic corpus, similar in this respect to their counterparts in the English corpus. The number of types is 4, which have a total frequency of 196. The type-token ratio in this category is 0.02. The four alternative types are "'aw" (the top frequency connective with a frequency of 152 representing 77.6% of total alternative tokens), "'am" (the next most frequent connective, having 25 occurrences that represent 12.8% of total alternatives), "'immā" (occurring as the core of the correlative connective "'immā ... wa 'immā" or "'immā ... 'aw", and having a frequency of 15 occurrences that represents 7.7% of total alternatives) and "sawa'an" (as a core of the correlative connective "sawā'an ... 'aw"; it is the least frequent alternative connective, having only 4 occurrences that represent 2% of total alternatives). The rank distribution is given in Table (8.16) (see also App. 47). Appendices (48-49) display the

Rank	Tokens	Types
1:	152	1
2:	25	1
3:	15	1
4:	4	1

Number of "Alternative" connective tokens: 196
Number of "Alternative" connective types : 4

Table 8.16 Rank distribution of alternative connectives in the Arabic corpus

distribution of alternative connectives within the entire corpus and within the connective tokens.

2. Repetitiveness

The following indices represent aspects of repetitiveness of alternative connectives in the corpus.

a. Alternative connective occurrence rate is approximately 87. This index suggests that alternative connectives occur at the rate of one in every 87 successive connectives in the corpus.

b. Type occurrence rate in this category is 49. This means that new types are encountered at the rate of one in every 49 successive alternative tokens.

c. The general repeat rate in this category is (13×10^{-5}). This means that the probability that two successive occurrences of connectives - with or without a separating gap - are alternatives is very low: 13 in 100,000.

d. The alternative system repeat rate is (83×10^{-6}). This refers to the probability that two successive occurrences of connectives are the same alternative connective.

e. The gap distribution shows that the average gap that separates two alternative connectives is 1,300 words. The shortest gap comprises one word only and has a frequency of 2. The largest gap is 13,364 and has a frequency of one. The highest frequency gap occurs only 3 and comprises 6 words.

3. Growth

Two calculations of growth of alternative connectives are computed. The results are given in Tables (8.17-18) and displayed in Appendices (50-51).

Tuldava's (token) richness index:	4.003811
Growth rate of "alternative" tokens:	-5.002525
Tuldava's (type) richness index:	0.007349
Growth rate of "alternative" types:	9.091329

Actual Text Tokens	Actual Alt.Tokens	Expected Alt.Tokens	Actual Alt.Types	Expected Alt.Types
5000	5	6	3	3
10000	13	11	3	3
15000	18	16	3	3
20000	20	20	4	3
25000	25	25	4	3
30000	36	29	4	3
35000	40	33	4	3
40000	46	37	4	3
45000	47	41	4	3
50000	49	45	4	3
55000	54	49	4	3
60000	55	53	4	3
65000	58	56	4	3
70000	63	60	4	3
75000	65	64	4	3
80000	68	67	4	3
85000	70	71	4	3
90000	73	75	4	3
95000	73	78	4	3
100000	73	82	4	3
105000	76	85	4	3
110000	76	89	4	3
115000	79	92	4	3
120000	85	96	4	3
125000	88	99	4	3
130000	91	102	4	3
135000	95	106	4	4
140000	98	109	4	4
145000	101	113	4	4
150000	104	116	4	4
155000	113	119	4	4
160000	116	123	4	4
165000	122	126	4	4
170000	127	129	4	4
175000	135	132	4	4
180000	139	136	4	4
185000	141	139	4	4
190000	143	142	4	4
195000	147	145	4	4
200000	161	149	4	4
205000	165	152	4	4
210000	168	155	4	4
215000	172	158	4	4

220000	175	161	4	4
225000	175	165	4	4
230000	177	168	4	4
235000	178	171	4	4
240000	180	174	4	4
245000	184	177	4	4
250000	188	180	4	4
255000	194	184	4	4
256450	196	184	4	4

Extrapolated Text Tokens	Expected Alt. Tokens	Expected Alt. Types
500000	330	4
750000	470	4
1000000	605	4

Table 8.17 Global growth of alternative connectives in the Arabic Corpus

a.Global Growth: This measure computes the growth of alternative types and tokens with intervals of 5,000 words of running text each in the corpus. The first interval contains 5 alternative tokens, which comprise 3 types. The expected number of tokens is 6. The highest number of alternative tokens within any interval is 14 which occur through interval 40. Extrapolation on the basis of the growth rate using Tuldava's formula gives the following results: an extrapolated corpus of 500,000 words of running text is expected to contain 330 alternative tokens but no increase in types; extrapolation of a million word corpus is expected to contain 605 alternative tokens, but the same 4 types. Tuldava's richness index of alternative tokens in relation to text tokens is 4.003811 and the growth rate of tokens is -5.002525. Tuldava's richness index of alternative types in relation to text tokens is 0.007349 and the growth rate of types is 9.091329.

Tuldava's richness index:	2.096689
"Alternative" Connective growth rate:	-3.062448

Actual Connective Tokens	Actual "Alternative" Connective Tokens	Expected "Alternative" Connective Tokens
500	5	8
1000	18	14
1500	23	21
2000	36	27
2500	41	33
3000	47	39
3500	50	45
4000	54	51
4500	58	57
5000	64	62
5500	68	68
6000	73	74
6500	73	79
7000	76	85
7500	82	90
8000	89	96
8500	91	101
9000	98	107
9500	103	112
10000	111	118
10500	118	123
11000	126	128
11500	135	134
12000	139	139
12500	142	144
13000	147	150
13500	162	155
14000	168	160
14500	175	165
15000	175	171
15500	178	176
16000	181	181
16500	188	186
16995	196	191

Extrapolated Connective Tokens	Expected "Alternative" Connective Tokens
20000	222
30000	323
50000	517
100000	983

Table 8.18 Local growth of alternative connectives in the Arabic corpus

b. Local Growth: This measure computes the growth of alternative connectives within intervals of 500 connective tokens. The first 500 connective tokens in the corpus contain only 5 connectives whereas the expected number, using Tuldava's calculation, is 8. The next 500 contains 13, bringing the number of alternatives to 18. An extrapolated number of 20,000 connective tokens is expected to give 221 alternatives. An extrapolation of 100,000 connective tokens is estimated to yield 983. Tuldava's richness index (of alternatives in relation to overall connective tokens) is 2.096689 and the growth rate of alternative tokens is -3.062448.

8.4 Comparative Connectives

8.4.1 Comparative Connectives in English

8.4.1.1 Global Profile

1. Frequency

The observed number of connectives of comparison in the English corpus is 15 types that have a total frequency of 441. The type-token ratio within this category is 0.034. The rank distribution (see Table 8.19) has 9 ranks. The first rank is occupied by the most frequent connective: "as", which has a frequency of 237 corresponding to approximately 54% of total comparative tokens. The next most frequent connective is "than", which has a frequency of 105 representing 24% of total comparatives (see the frequency list in Appendices 52-53). The lowest rank is occupied by two hapaxes. The diagrams in Appendices (54-55) plot the distribution of comparative connectives in relation to the text tokens and to connective tokens.

Rank	Tokens	Types
1:	237	1
2:	105	1
3:	33	1
4:	14	2
5:	7	3
6:	5	1
7:	3	2
8:	2	2
9:	1	2

Number of "Comparative" connective tokens: 441
Number of "Comparative" connective types : 15

Table 8.19 Rank distribution of comparative connectives in the English corpus

2. Repetitiveness

The following indices represent some aspects of repetitiveness of the connectives of this category.

a. Comparative connective occurrence rate is approximately 22. This indicates that one comparative connective is expected to appear in every 22 successive occurrences of connective tokens.

b. Type occurrence rate in this category is 29. This means that a new comparative type is encountered in every 29 occurrences of comparative tokens in the corpus.

c. The general repeat rate index is 0.002. This represents the probability that two successive occurrences of connectives - with or without a gap separating them - are any comparative connectives.

d. The comparative system repeat rate is (75×10^{-5}). This

indicates that the probability that two successive connectives in the corpus turn out to be the same comparative connective is 75 in 100,000.

e. Gap distribution of comparative connectives reveals that the average gap between two occurrences of comparatives is 579 words. The shortest gap comprises 2 words and has a frequency of 2. The largest gap consists of 2,888 words and has one occurrence. The most frequent gap has a frequency of 5 and comprises 43 words.

3. Growth

Two calculations of growth of comparative connectives are made: global and local. The results are given in Tables (8.20-21) and displayed in the diagrams of Appendices (56-57).

Tuldava's (token) richness index:	4.057289
Growth rate of "comparative" tokens:	-5.026623
Tuldava's (type) richness index:	1.011648
Growth rate of "comparative" types:	-7.059404

Actual Text Tokens	Actual Com.Tokens	Expected Com.Tokens	Actual Com.Types	Expected Com.Types
5000	12	11	4	4
10000	24	21	5	5
15000	30	31	7	6
20000	37	41	7	7
25000	51	50	8	8
30000	57	59	8	8
35000	66	68	10	8
40000	73	78	10	9
45000	85	86	10	9
50000	96	95	11	10
55000	109	104	11	10
60000	116	113	11	10
65000	126	122	12	10
70000	135	130	13	11
75000	142	139	13	11
80000	149	148	13	11

85000	163	156	13	11
90000	171	165	13	12
95000	178	173	13	12
100000	186	182	14	12
105000	190	190	15	12
110000	198	198	15	13
115000	204	207	15	13
120000	211	215	15	13
125000	221	224	15	13
130000	226	232	15	13
135000	240	240	15	14
140000	256	248	15	14
145000	264	257	15	14
150000	272	265	15	14
155000	281	273	15	14
160000	289	281	15	14
165000	298	290	15	14
170000	305	298	15	15
175000	309	306	15	15
180000	314	314	15	15
185000	320	322	15	15
190000	329	330	15	15
195000	339	338	15	15
200000	347	346	15	15
205000	351	355	15	16
210000	356	363	15	16
215000	360	371	15	16
220000	368	379	15	16
225000	377	387	15	16
230000	391	395	15	16
235000	395	403	15	16
240000	410	411	15	17
245000	417	419	15	17
250000	427	427	15	17
255000	438	435	15	17
256560	441	437	15	17

Extrapolated Text Tokens	Expected Com. Tokens	Expected Com. Types
500000	816	21
750000	1193	25
1000000	1562	27

Table 8.20 Global growth of comparative connectives in the English corpus

a. Global Growth: This measure monitors the growth of comparative connectives within intervals of 5,000 words of running text (the corpus is thus divided into 52 intervals). Tuldava's formula is also applied to compute the expected growth and for extrapolation purposes. The first 5,000 words contains 12 comparatives that comprise 4 types. The expected figures are close: 11 tokens and 4 types. The largest number of comparative tokens observed at any interval is 15 (at interval 48). An extrapolated corpus of 500,000 running text is expected to include 816 comparative tokens and 21 types. An extrapolated corpus of a million words is expected to comprise 1,562 tokens and 27 types. Tuldava's richness index of comparative tokens in relation to the entire text tokens is 4.057289 and the growth rate of tokens within text intervals is -5.026623. Tuldava's richness index of types in relation to the corpus is 1.011648 and the growth rate of types is -7.059404.

b. Local Growth: This measure monitors the growth of comparative connective tokens within intervals of 500 successive connective tokens in the corpus. The first 500 connective tokens contain 27 comparatives. The next 500 comprise another 27, raising the number of comparatives to 54. The expected number in this interval is 24, which raises the number of 51. Extrapolation of 10,000 connective tokens is expected to contain 447 comparative connectives. An extrapolated number of 100,000 connective tokens is estimated to yield 3,967 comparatives. Tuldava's richness index (referring to comparative tokens in relation to connective tokens) is 2.012836 and the growth rate in the intervals is -2.056534.

Tuldava's richness index:	2.012836
"Comparative" Connective growth rate:	-2.056534

Actual Connective Tokens	Actual "Comparative" Connective Tokens	Expected "Comparative" Connective Tokens
500	27	27
1000	54	51
1500	73	75
2000	99	98
2500	126	121
3000	143	144
3500	170	166
4000	188	189
4500	205	211
5000	224	233
5500	261	254
6000	282	276
6500	301	298
7000	316	319
7500	339	341
8000	357	362
8500	377	383
9000	406	405
9500	433	426
9596	441	430

Extrapolated Connective Tokens	Expected "Comparative" Connective Tokens
10000	447
15000	656
20000	861
30000	1263
50000	2052
100000	3967

Table 8.21 Local growth of comparative connectives in the English corpus

8.4.1.2 Categories of Comparative Connectives in English

Comparative connectives are divided into two main functional categories: degree and similarity. The statistical distribution of those two categories is shown in Table (8.22) (see also the frequency lists for each category in Appendices (58A-B). Figure (8.7) displays the proportion of each of these categories in terms of comparative tokens.

According to Table (8.22), the total number of observed types in both categories are 20. This creates an increase of 5 types over global comparative connectives. This increase is created by the

Category	Tokens	% Tokens	Types	% Types
Degree	208	47.17	9	45
Similarity	233	52.83	11	55
Total	441	100.00	20	100.00

Table 8.22 Distribution of categories of comparative connectives in the English corpus

connective "as", which, although as a core is counted as 1, represents at this level of categorisation, 5 connectives:

a) as a one-word connective expressing degree,

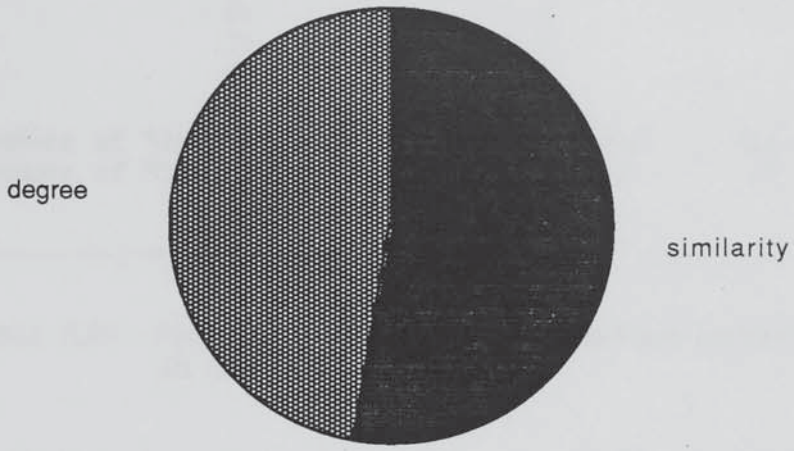
b) as a one-word connective expressing similarity,

c) as a correlate connective expressing degree "as ... as",

d) as a correlate connective expressing similarity "just as..so",

e) as a compound (multi-word) connective expressing similarity in "as if".

Fig. 8.7 Distribution of Categories of Comparative Connectives in the English Corpus



Comparative connectives of "similarity" represent the larger of the two categories, both in terms of types and tokens. It comprises a set that corresponds to approximately 53% of comparative tokens and 55% of types.

8.4.2 Comparative Connectives in Arabic

8.4.2.1 Global Profile

1. Frequency

Connectives of comparison observed in the Arabic corpus has 12 types and a frequency of 266 (see Table 8.23 and Appendices 59-60). The distribution consists of 8 ranks. The top rank is occupied by the connective "kamā" with a frequency of 121 that represents 45.5% of total comparative tokens. The next rank is reserved for the

Rank	Tokens	Types
1:	121	1
2:	73	1
3:	24	1
4:	17	2
5:	4	1
6:	3	1
7:	2	2
8:	1	3

Number of "Comparative" connective tokens: 266
Number of "Comparative" connective types : 12

Table 8.23 Rank distribution of comparative connectives in the Arabic corpus

connective "ka'anna", which has a frequency of 73 representing 27% of total comparatives. The lowest rank is filled by 2 hapaxes "ka'annamā" and "<alā hāḏā al-wajhi" (the word "wajhi" is considered the core element of this compound connective). The type-token ratio in this category is 0.045. The diagrams in Appendices (61-62) display the distribution of comparatives within the corpus and within connective tokens.

2. Repetitiveness

The following indices represent various aspects of repetitiveness of comparative connectives within the corpus.

a. Occurrence rate of comparative connectives is 64. This indicates that one comparative connective is expected in each set of 64 successive connectives in the corpus.

b. Type occurrence rate is 22. That is, a new comparative type is expected to appear in each 22 successive occurrences of

comparative tokens.

c. The general repeat rate index is (24×10^{-5}). This refers to the probability that two successive connectives - with or without a gap separating them - turn out to be any comparative connectives.

d. The comparative system repeat rate is (7×10^{-5}). This represents the probability that two successive connectives are the same comparative type.

e. Calculation of gap distribution shows that the average gap length is 959 words. The shortest gap length consists of one word and has one occurrence, while the longest gap length comprises 6,818 and has one occurrence. The most frequent gap lengths are five and seven words, each with a frequency of four.

3. Growth

Two calculations of growth are produced for comparative connectives: global and local. In addition, Tuldava's formula has been applied to compute the expected growth and for extrapolation purposes. Results of these computations are given in Tables (8.24-25) and displayed in Appendices (63-64).

a. Global Growth: This measure computes the growth of comparative connectives within intervals of 5,000 words of running text. The first interval includes 12 comparative tokens, which consist of 7 types. The expected numbers at this interval is 9 tokens and 6 types. The second interval shows a growth of 6 tokens raising the number of tokens to 18, and one type, raising the number to 8. The expected growth is 7 tokens and no increase in types. The largest growth in terms of tokens at any one interval is 12 at interval one, followed by 10 at intervals 29 and 35. The smallest growth is one

---

Tuldava's (token) richness index:	3.060098
Growth rate of "comparative" tokens:	-4.085748
Tuldava's (type) richness index:	0.693378
Growth rate of "comparative" types:	11.007019

Actual Text Tokens	Actual Com.Tokens	Expected Com.Tokens	Actual Com.Types	Expected Com.Types
5000	12	9	7	6
10000	18	16	8	6
15000	23	23	8	7
20000	30	29	8	7
25000	36	35	8	7
30000	39	41	8	8
35000	44	47	8	8
40000	49	52	8	8
45000	53	58	8	8
50000	58	63	8	8
55000	67	68	8	8
60000	70	74	8	9
65000	76	79	8	9
70000	83	84	9	9
75000	84	89	9	9
80000	88	94	9	9
85000	93	99	9	9
90000	97	104	9	9
95000	101	109	9	9
100000	109	114	9	9
105000	113	119	9	9
110000	116	123	9	9
115000	120	128	9	10
120000	127	133	9	10
125000	135	137	9	10
130000	137	142	9	10
135000	141	147	9	10
140000	149	151	10	10
145000	159	156	10	10
150000	164	161	10	10
155000	167	165	11	10
160000	172	170	11	10
165000	179	174	11	10
170000	183	179	11	10
175000	193	183	12	10
180000	200	188	12	10
185000	203	192	12	10
190000	209	197	12	10
195000	212	201	12	10
200000	216	205	12	10
205000	220	210	12	10
210000	222	214	12	10
215000	230	219	12	11

220000	235	223	12	11
225000	241	227	12	11
230000	245	232	12	11
235000	250	236	12	11
240000	253	240	12	11
245000	253	244	12	11
250000	259	249	12	11
255000	264	253	12	11
256450	266	254	12	11

Extrapolated Text Tokens	Expected Com. Tokens	Expected Com. Types
500000	452	12
750000	641	13
1000000	823	13

Table 8.24 Global growth of comparative connectives in the Arabic corpus

token at interval 15. Types are saturated at interval 35.

An extrapolated corpus of 500,000 words is expected to include 452 comparative tokens while the number of types is expected to remain 12. An extrapolated corpus of a million words is expected to include 823 comparative tokens, but the number of types is expected to be 13, i.e. an increase of one type. Tuldava's richness index of comparative tokens within the corpus is 3.060098, and of comparative types within the corpus is 0.693378. Growth rate of comparative tokens within the intervals is -4.085748 and growth rate of types within the intervals is 11.007019.

b. Local Growth: This measure computes the growth of comparative tokens within intervals of 500 successive connective tokens each (the corpus is thus divided into 34 intervals). The first 500 connective tokens contains 14 comparative tokens; the expected

Tuldava's richness index:	2.028133
"Comparative" Connective growth rate:	-3.010706

Actual Connective Tokens	Actual "Comparative" Connective Tokens	Expected "Comparative" Connective Tokens
500	14	12
1000	23	22
1500	32	31
2000	39	40
2500	47	48
3000	52	56
3500	60	65
4000	69	73
4500	76	80
5000	83	88
5500	88	96
6000	97	104
6500	109	111
7000	113	119
7500	125	126
8000	135	134
8500	141	141
9000	150	148
9500	162	155
10000	167	163
10500	173	170
11000	181	177
11500	193	184
12000	200	191
12500	207	198
13000	212	205
13500	218	212
14000	222	219
14500	232	226
15000	241	233
15500	250	240
16000	253	247
16500	259	253
16995	266	260

Extrapolated Connective Tokens	Expected "Comparative" Connective Tokens
20000	301
30000	431
50000	681
100000	1268

Table 8.25 Local growth of comparative connectives in the Arabic corpus

number is 12. The next interval monitors a growth of 9 connectives, raising the number of tokens to 22. The largest growth at any interval is 14 at interval one, followed by 12 at intervals 13, 15 and 23. An extrapolated number of 20,000 connective tokens is expected to include 301 comparative tokens. Extrapolation of 100,000 connectives is estimated to incorporate 1,268 comparatives. Tuldava's richness index of comparatives within the overall connective tokens is 2.028133. The growth rate of comparatives within the connective intervals is -3.010706.

8.4.2.2 Categories of Comparative Connectives in Arabic

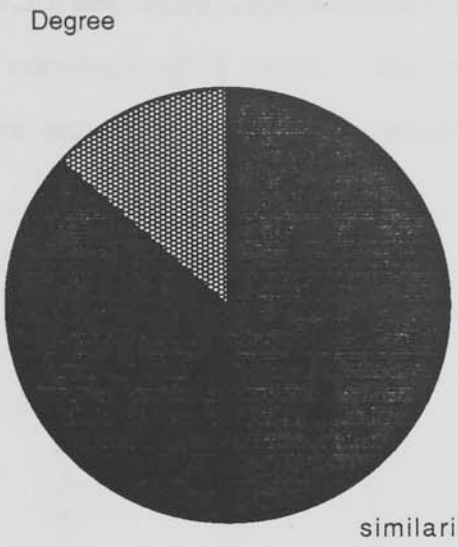
The distribution of the two categories of comparative connectives is shown in Table (8.26). The proportion of each category in terms of comparative tokens is displayed in Figure (8.8). The total number of observed types in the two categories is 15. This represents an increase of 3 types over global comparative types and

Category	Tokens	% Tokens	Types	% Types
Degree	35	13.16	4	26.67
Similarity	231	86.84	11	73.33
Total	266	100.00	15	100.00

Table8.26 Distribution of categories of comparative connectives in the Arabic corpus

is effected by the use of the connective "kamā". Although at the global computation of comparative connectives, this connective is calculated as one core, it stands at this detailed level of categorisation as a core for 3 connectives: as a one-word connective, as a compound connective in "kamā law", as a correlate

Fig. 8.8 Distribution of Categories of Comparative Connectives in the Arabic Corpus



connective in "kamā ... fa". In addition, the core "qadri" occurs in two connectives: a compound "bi-qadri mā" and correlate "<alā qadri".

The comparative category of similarity is distinctly larger than that of degree, both in terms of tokens and types. It comprises approximately 87% of comparative tokens and 73% of types, which indicates a more extensive textual role than that of the other comparative category (see frequency lists in Appendices 65A-B).

8.5 Reformulatory Connectives

8.5.1 Reformulatory Connectives in English

8.5.1.1 Global Profile

1. Frequency

Observation of connectives of this category in the English corpus

indicates that they comprise 18 types that have a total frequency of 115. The type-token ratio in this category is 0.157. The rank distribution (Table 8.27 see also alphabetical and frequency lists in Appendices 66-67) consists of 9 ranks. The top rank is occupied by the connective "for example" with a frequency of 36 representing

Rank	Tokens	Types
1:	36	1
2:	12	3
3:	11	1
4:	8	1
5:	5	1
6:	4	1
7:	3	1
8:	2	3
9:	1	6

Number of "Reformulatory" connective tokens: 115
Number of "Reformulatory" connective types : 18

Table 8.27 Rank distribution of reformulatory connectives in the English corpus

about 31% of total reformulatory tokens. The next rank is occupied by 3 connectives, each having a frequency of 12 which is equivalent to 10.4% of total reformulatory tokens. These are "for instance", "rather", "in other words". Thus the total frequency of the first two ranks corresponds to more than 62% of total reformulatory occurrences. The lowest rank comprises 6 hapaxes: "more accurately", "more pointedly", "in simple terms", "briefly", "in brief" and "in all".

2. Repetitiveness

Indices of repetitiveness of connectives of this category have been computed as follows:

a. The reformulatory occurrence rate is slightly more than 83. This means that, on average, one connective of reformulation appears in each set of 83 successive connective tokens.

b. Type occurrence rate within this category is 6. That is, a new type is encountered in every 6 successive occurrences of reformulatory connective tokens.

c. General reformulatory repeat rate, referring to the probability that 2 successive connectives - with or without a gap separating them - are reformulatory, is (14×10^{-5}), i.e. 14 in 100,000.

d. Reformulatory system repeat rate, referring to the probability that two successive connectives - with or without a gap separating them - are the same reformulatory connective, is (22×10^{-6}), i.e. 22 in a million.

e. Results of the gap measurement indicate that the average distance separating two sites each occupied by a reformulatory connective is 2,209 words. All distance lengths have one occurrence each. The shortest gap length is 46 words and the longest 16,214.

3. Growth

Results of the calculation of growth are given in Tables (8.28-29) and represented visually in Appendices (70-71).

a. Global Growth: This measure examines the growth of connective types and tokens within successive intervals, each of 5,000 words of running text. The first interval contains 3 tokens, each a different type. The expected numbers, using Tuldava's formula is 4 tokens and 3 types. An examination of growth over the intervals

Tuldava's (token) richness index:	4.038148
Growth rate of "reformulatory" tokens:	-5.066202
Tuldava's (type) richness index:	1.043357
Growth rate of "reformulatory" types:	-5.008157

Actual Text Tokens	Actual Ref.Tokens	Expected Ref.Tokens	Actual Ref.Types	Expected Ref.Types
5000	3	4	3	3
10000	9	7	5	4
15000	11	10	5	5
20000	15	12	7	6
25000	19	15	8	6
30000	21	17	9	7
35000	21	20	9	7
40000	21	22	9	8
45000	24	25	9	8
50000	27	27	9	8
55000	29	29	9	9
60000	29	31	9	9
65000	32	34	10	9
70000	34	36	10	10
75000	37	38	10	10
80000	39	40	11	10
85000	42	42	11	11
90000	44	45	11	11
95000	45	47	11	11
100000	47	49	12	11
105000	48	51	12	12
110000	50	53	12	12
115000	54	55	12	12
120000	54	57	12	12
125000	58	59	14	13
130000	58	61	14	13
135000	60	63	14	13
140000	62	65	14	13
145000	65	67	14	13
150000	66	69	14	14
155000	69	71	14	14
160000	72	73	14	14
165000	77	75	15	14
170000	80	77	15	14
175000	82	79	15	14
180000	84	81	15	15
185000	85	83	15	15
190000	87	85	15	15
195000	90	87	15	15
200000	92	89	16	15
205000	94	91	16	15
210000	95	92	16	16
215000	98	94	16	16
220000	101	96	16	16
225000	102	98	16	16

230000	103	100	16	16
235000	104	102	16	16
240000	106	104	17	17
245000	112	106	17	17
250000	114	107	18	17
255000	115	109	18	17
256560	115	110	18	17

Extrapolated Text Tokens	Expected Ref. Tokens	Expected Ref. Types
500000	196	23
750000	278	27
1000000	358	31

Table 8.28 Global growth of reformulatory connectives in the English corpus

shows that it stops during intervals 7 and 8, then resumes at a fairly similar rate. It also stops at these intervals: 12, 24, 26 and at the final, shorter interval. Types show first signs of saturation at interval 7 where growth stops until a new type appears in interval 13. An extrapolated corpus of 500,000 words is expected to contain 196 tokens and 23 types. An extrapolated corpus of a million words is expected to include 358 tokens and 31 types.

b. Local Growth: This measure computes the growth of reformulatory connectives within intervals of 500 tokens each. The first interval includes 9 connectives of this category, which is the same as the expected number. Growth fluctuates within the intervals from 11, the highest level of growth, at interval 2 to one, the lowest level of growth at interval 3. Note that at the end of the final interval the expected number of connectives is smaller than the observed. An extrapolated number of 10,000 connective tokens is expected to include 111 reformulatory tokens, a still lower number

Tuldava's richness index:	2.034799
"Reformulatory" Connective growth rate:	-3.031883

Actual Connective Tokens	Actual "Reformu." Connective Tokens	Expected "Reformu." Connective Tokens
500	9	9
1000	20	16
1500	21	22
2000	28	28
2500	32	34
3000	37	40
3500	44	45
4000	47	51
4500	54	56
5000	58	61
5500	65	66
6000	69	72
6500	78	77
7000	84	82
7500	90	87
8000	96	92
8500	102	97
9000	106	101
9500	114	106
9596	115	107

Extrapolated Connective Tokens	Expected "Reformulatory" Connective Tokens
10000	111
15000	158
20000	202
30000	287
50000	449
100000	824

Table 8.29 Local growth of reformulatory connectives in the English Corpus

than the observed one in the corpus. This is mainly due to the high fluctuation of growth, which causes the observed figures to deviate from the expected ones. An extrapolated number of 100,000 connective tokens is expected to yield 824 reformulatory tokens.

8.5.1.2 Categories of Reformulatory Connectives in English

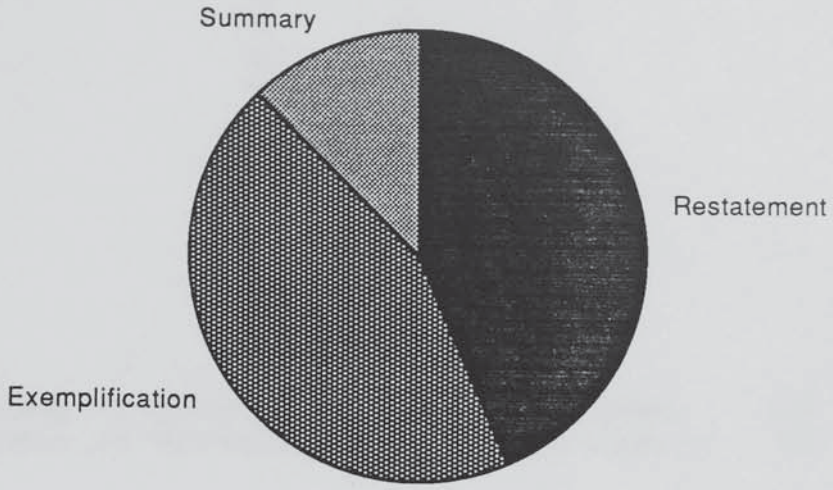
Distribution of categories of reformulatory connectives are given in Table (8.30) (see also the frequency lists in Appendices 72A-C). Figure (8.9) displays the share of each category in token mass. The total number of types in the three categories of reformulatory connectives is 19, an increase of one type over global reformulatory types. This is because the connective "rather" occurs as a core for another connective "or rather".

Category	Tokens	% Tokens	Types	% Types
Restatement	50	43.48	12	63.16
Exemplification	51	44.35	3	15.79
Summary	14	12.17	4	21.05
Total	115	100.00	19	100.00

Table 8.30 Distribution of the categories of reformulatory connectives in the English corpus

Inspection of the Table shows that the categories of restatement and exemplification have similar relative frequency but differ in the number of types that each comprises. Thus while the type-token ratio is 0.24 in the restatement category, it is 0.06 in the exemplification category. The smallest category is that of summary. It has a relative frequency of only 12% among reformulatory connectives.

Fig. 8.9 Distribution of the categories of Reformulatory connectives in the English Corpus



8.5.2 Reformulatory Connectives in Arabic

8.5.2.1 Global Profile

1. Frequency

Calculation of reformulatory connectives in the Arabic corpus indicates that there are 27 types that have a total frequency of 239. The rank distribution (Table 8.31; see also frequency lists in Appendices 73-74) is composed of 12 ranks. The top rank is taken up by the connective "'ay" [that is, in other words], which has a frequency of 80 representing 33.5% of total reformulatory tokens. The second rank is reserved for the connective "maṯalan" [for example] and has a frequency of 41 corresponding to 17.2% of reformulatory tokens. Thus these two top ranks achieve more than half the reformulatory token coverage in the corpus. The lowest

Rank	Tokens	Types
1:	80	1
2:	41	1
3:	26	1
4:	19	1
5:	13	1
6:	9	1
7:	8	1
8:	7	1
9:	5	2
10:	4	2
11:	2	3
12:	1	12

Number of "Reformulatory" connective tokens: 239
Number of "Reformulatory" connective types : 27

Table 8.31 Rank distribution of reformulatory connectives in the Arabic corpus

rank is occupied by 13 hapaxes that, while representing 48% of types, is equivalent to only 5.4% of tokens. The type-token ratio in this category is 0.113. Appendices (75-76) display the distribution of reformulatory tokens in the corpus and within connective tokens respectively.

2. Repetitiveness

Indices of repetitiveness have been computed as follows:

a. Reformulatory occurrence rate is 71, indicating that a connective signalling reformulation is encountered in every 71 successive connectives.

b. Type occurrence rate within this category is 9, which suggests that a new type appears in an average of 9 successive reformulatory tokens.

c. General repeat rate of reformulatory connectives is (2×10^{-4}). This means that the probability that two successive connective tokens - with or without gaps separating them - turn out to be both reformulatory connectives is 2 in 10,000.

d. Reformulatory system repeat rate is (33×10^{-6}). This refers to the probability that two successive connectives are the same reformulatory type.

e. Results of gap measurements show that the average gap length intervening between two reformulatory tokens is 1,067 words. The shortest gap length comprises 10 words and has one occurrence. The longest gap length consists of 7,622 words and a frequency of one. The most frequent gaps are 81 and 256 words, each with a frequency of 3.

3. Growth

Results of the calculation of global and local growth of reformulatory connectives are given in Tables (8.32-33) and displayed in Appendices (77-78).

Tuldava's (token) richness index:	7.026972
Growth rate of "reformulatory" tokens:	-7.014547
Tuldava's (type) richness index:	1.005797
Growth rate of "reformulatory" types:	-5.020216

Actual Text Tokens	Actual Ref.Tokens	Expected Ref.Tokens	Actual Ref.Types	Expected Ref.Types
5000	2	4	2	4
10000	10	9	6	6
15000	17	13	11	7
20000	22	18	11	8
25000	33	23	13	9
30000	34	27	13	9
35000	42	32	13	10
40000	46	37	13	11
45000	48	42	13	11

50000	50	46	13	12
55000	51	51	13	13
60000	57	56	14	13
65000	62	60	14	14
70000	70	65	14	14
75000	74	70	15	15
80000	75	75	15	15
85000	79	79	15	16
90000	82	84	16	16
95000	84	89	16	16
100000	86	94	16	17
105000	91	98	16	17
110000	94	103	16	18
115000	101	108	17	18
120000	106	113	18	18
125000	108	118	18	19
130000	118	122	19	19
135000	119	127	19	20
140000	123	132	20	20
145000	126	137	21	20
150000	131	141	22	21
155000	135	146	22	21
160000	148	151	22	21
165000	155	156	22	22
170000	158	161	22	22
175000	166	165	22	22
180000	176	170	22	22
185000	176	175	22	23
190000	179	180	22	23
195000	190	184	23	23
200000	193	189	24	24
205000	198	194	24	24
210000	202	199	24	24
215000	205	204	25	25
220000	213	208	25	25
225000	214	213	25	25
230000	217	218	26	25
235000	222	223	26	26
240000	226	228	27	26
245000	230	232	27	26
250000	231	237	27	26
255000	237	242	27	27
256450	239	243	27	27

Extrapolated Text Tokens	Expected Ref. Tokens	Expected Ref. Types
500000	478	37
750000	720	46
1000000	963	53

Table 8.32 Global growth of reformulatory connectives in the Arabic corpus

Tuldava's richness index:	5.059982
"Reformulatory" Connective growth rate:	-4.097773

Actual Connective Tokens	Actual "Reformu." Connective Tokens	Expected "Reformu." Connective Tokens
500	3	5
1000	14	12
1500	24	19
2000	34	25
2500	44	32
3000	48	39
3500	50	47
4000	55	54
4500	62	61
5000	72	68
5500	75	76
6000	82	83
6500	86	91
7000	93	98
7500	104	105
8000	109	113
8500	118	120
9000	123	128
9500	131	136
10000	135	143
10500	148	151
11000	158	158
11500	164	166
12000	176	174
12500	177	181
13000	190	189
13500	197	197
14000	202	204
14500	213	212
15000	216	220
15500	221	228
16000	227	235
16500	230	243
16995	239	251

Extrapolated Connective Tokens	Expected "Reformulatory" Connective Tokens
20000	298
30000	457
50000	781
100000	1614

Table 8.33 Local growth of reformulatory connectives in the Arabic corpus

a. Global Growth: This measure monitors the growth of reformulatory connective tokens and types within interval of text, each 5,000 words of running text. The first interval contains only 2 tokens both different types. The expected number is 4. Growth of tokens fluctuates between 13, which is the highest level, at interval 32, to a temporary stoppage of growth at interval 37. Growth of types is fast during the first 3 intervals but then slumps to a minimum level, which is an indication of early saturation. An extrapolated corpus of 500,000 words is expected to contain 478 reformulatory tokens and 37 types. An extrapolated corpus of a million words is estimated to include 963 reformulatory tokens and 53 types.

b. Local Growth: This monitors the growth of reformulatory tokens within intervals of 500 successive connectives. The first interval includes 3 tokens while the expected number is 5. Growth within the intervals fluctuates between 13, the highest level, at intervals 21 and 26, to one at interval 25. An extrapolated number of 20,000 connectives is expected to include 298 reformulatory tokens. An extrapolation of 100,000 connectives is expected to include 1,614.

8.5.2.2 Categories of Reformulatory Connectives in Arabic

The distribution of the three categories of reformulatory connectives in the Arabic corpus is summarised in Table (8.34). The share that each category has in the reformulatory token mass is displayed in Figure (8.10). More detailed frequency lists are given in Appendices (79A-C). The number of types is 29, having an

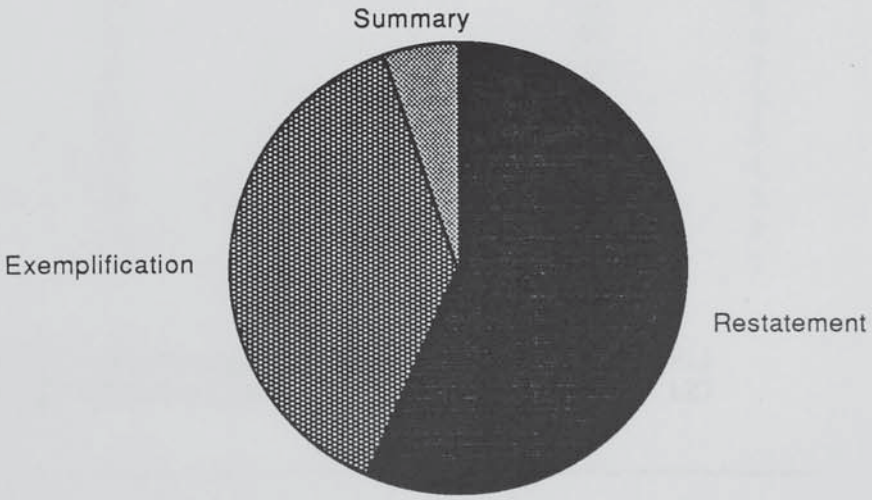
Category	Tokens	% Tokens	Types	% Types
Restatement	136	56.90	15	51.73
Exemplification	91	38.08	10	34.48
Summary	12	5.02	4	13.79
Total	239	100.00	29	100.00

Table 8.34 Distribution of the categories of reformulatory connectives in the Arabic corpus

increase of 2 types over global reformulatory types. This is because the connectives "ya<nī" and "maṯalan" occur as cores in the compound connectives "ya<nī hāḏā" and "li-naḍrib maṯalan" respectively.

The largest category of reformulatory connectives is that of restatement, which has a relative frequency of about 57%. It also comprises the biggest number of types. The next category in size is

Fig. 8.10 Distribution of the categories of Reformulatory connectives in the Arabic Corpus



that of exemplification, which has a relative frequency of 38%. Connectives of summary constitute the smallest category both in terms of tokens and types.

8.6 Orientative Connectives

8.6.1 Orientative Connectives in English

8.6.1.1 Global Profile

1. Frequency

This category incorporates the largest number of types in comparison with other categories. There are 127 types that have a total frequency of 1,011, which gives a type-token ratio of 0.126. The rank distribution (Table 8.35, see also frequency lists in Appendices 80-81) consists of 26 ranks. The top rank is occupied by the most frequent of the connectives of this category: "as" (used

Rank	Tokens	Types	Rank	Tokens	Types
1:	190	1	14:	15	1
2:	108	1	15:	12	2
3:	61	1	16:	11	3
4:	54	1	17:	10	2
5:	36	1	18:	9	2
6:	31	1	19:	8	2
7:	25	1	20:	7	5
8:	23	1	21:	6	9
9:	22	1	22:	5	7
10:	20	1	23:	4	7
11:	18	1	24:	3	11
12:	17	1	25:	2	16
13:	16	1	26:	1	47

Number of "Orientative" connective tokens: 1011
Number of "Orientative" connective types : 127

Table 8.35 Rank distribution of orientative connectives in the English corpus

for assigning authority). It has a frequency of 190 that corresponds to 18.85% of total orientative connective tokens. The next rank is reserved for the connective "indeed", occurring in the corpus with a frequency of 107 which is equivalent to 10.62% of total orientative occurrences. Other high frequency connectives, occupying the next few ranks, include: "certainly" (61 occurrences in a connective role, representing 6.05% of orientatives), "of course" (54 occurrences representing 5.36% of orientative tokens), "of course" (36 occurrences representing 3.57% of orientatives), and "actually" (functioning as a connective, with a frequency of 31 representing 3.08% of orientatives). The distributions of orientative tokens within the corpus and in relation to total connective tokens are displayed in Appendices (82-83) respectively.

The first 13 ranks, occupied by one connective each, correspond to approximately 10% of types but achieve a token coverage of more than 61%. On the other hand, the lowest rank which contains 48 hapaxes, stands for about 38% of types, though corresponding to less than 5% orientative token coverage. The type-token ratio of orientative connectives is 0.126.

2. Repetitiveness

Indices of repetitiveness have been computed as follows:

a. Orientative occurrence rate is 17. This means that one orientative connective appears within an average of 17 successive connectives in the corpus.

b. Type occurrence rate in this category is 8. In other words, a new orientative type is encountered in each 8 successive orientative tokens.

c. General repeat rate is 0.011. This refers to the probability that two successive connective tokens - with or without a gap separating them - are any orientatives.

d. Orientative system repeat rate is (67×10^{-5}). That is, the probability that two successive connectives turn out to be the same orientative type is 67 in 100,000.

e. Calculation of gaps separating occurrences of orientative connectives suggest that the average distance length is 252 words. The shortest distance is a 0 length (i.e. two orientative tokens follow each other immediately) and has a frequency of 9. The longest comprises 1,946 words and has a frequency of one. The most frequent distance is 54 words, which has a frequency of 14.

3. Growth

Results of the calculation of global and local growth are displayed in Tables (8.36-37) and represented visually in Appendices (84-85).

Tuldava's (token) richness index:	6.008863
Growth rate of "orientative" tokens:	-6.031174
Tuldava's (type) richness index:	1.048117
Growth rate of "orientative" types:	-4.015065

Actual Text Tokens	Actual Orien.Tokens	Expected Orien.Tokens	Actual Orien.Types	Expected Orien.Types
5000	15	17	11	14
10000	41	36	20	20
15000	54	55	23	26
20000	77	74	30	30
25000	96	94	38	35
30000	128	114	43	39
35000	139	134	46	42
40000	149	154	49	46
45000	167	175	52	49
50000	191	195	57	52
55000	206	215	61	55

60000	221	236	63	58
65000	240	257	67	61
70000	260	277	70	64
75000	292	298	71	67
80000	321	319	77	70
85000	340	340	78	72
90000	365	361	80	75
95000	384	382	81	77
100000	409	403	84	80
105000	437	424	88	82
110000	463	445	92	84
115000	489	466	95	87
120000	508	487	96	89
125000	532	509	96	91
130000	556	530	97	93
135000	568	551	98	95
140000	591	572	101	98
145000	616	594	103	100
150000	639	615	104	102
155000	665	637	106	104
160000	681	658	107	106
165000	706	679	107	108
170000	729	701	108	110
175000	750	722	110	112
180000	762	744	111	114
185000	785	766	113	116
190000	801	787	114	118
195000	816	809	116	119
200000	831	830	117	121
205000	844	852	119	123
210000	858	874	119	125
215000	870	895	119	127
220000	894	917	119	129
225000	907	939	120	130
230000	925	960	122	132
235000	943	982	125	134
240000	966	1004	126	136
245000	984	1026	126	137
250000	993	1047	127	139
255000	1009	1069	127	141
256560	1011	1076	127	141

Extrapolated Text Tokens	Expected Orien. Tokens	Expected Orien. Types
500000	2153	214
750000	3278	275
1000000	4417	329

Table 8.36 Global growth of orientative connectives in the English corpus

a. Global Growth: This measure traces the growth of orientative tokens and types within text intervals of 5,000 words each. The first interval contains 15 orientative tokens and 11 types, giving a high ratio of 0.73. The next interval contains 26 more tokens, raising the number to 41, and 9 new types. The ratio now drops to 0.48. Growth of tokens varies within the intervals from 32, which is observed at interval 6 and which represents the highest level of growth at any interval, to 9, its lowest level observed at interval 9. (As with other calculations of growth, interval 52 is excluded because of its small size.) Growth of new types fluctuates to a large extent and although signs of saturation start to appear from interval 28 onwards, new types continue to emerge in the following intervals. The 50th type percentile is observed at interval 12 (60,000 words); the 75th type percentile occurs within interval 21 (105,000 words) and the 90th type percentile occurs within interval 38 (190,000 words). An extrapolated corpus of 500,000 words is expected to include 2,153 orientative tokens and 214 types. An extrapolated corpus of 1,000,000 is expected to contain 4,417 tokens and 329 types.

b. Local Growth: This measure computes the growth of orientative tokens within intervals of 500 successive connectives in the corpus. The first interval contains 53 orientative tokens. In the next interval there is a growth of 51 tokens, raising the number to 104 (the expected number at this interval is 102). Growth varies between 71 at interval 12 to 39 at interval 5. An extrapolated number of 10,000 connective tokens is expected to contain 1,078 orientatives. An extrapolated number of 100,000 connectives is estimated to include 11,246.

Tuldava's richness index:	2.068483
"Orientative" Connective growth rate:	-2.047613

Actual Connective Tokens	Actual "Orientative" Connective Tokens	Expected "Orientative" Connective Tokens
500	53	50
1000	104	102
1500	149	154
2000	200	208
2500	239	261
3000	300	315
3500	362	369
4000	427	423
4500	490	477
5000	550	531
5500	595	585
6000	666	640
6500	717	695
7000	775	749
7500	820	804
8000	862	859
8500	909	914
9000	961	968
9500	1007	1023
9596	1011	1034

Extrapolated Connective Tokens	Expected "Orientative" Connective Tokens
10000	1078
15000	1631
20000	2187
30000	3305
50000	5559
100000	11246

Table 8.37 Local growth of orientative connectives in the English corpus

8.6.1.2 Categories of English Orientative Connectives

Distribution of categories of orientative connectives is given in Table (8.38) and displayed in Figure 8.11. More detailed results are given in the lists in Appendices (86A-B). The total number of observed types in the orientative categories is 129, an increase of two types over global orientative types. This is because the connective "as" and "crucially" occur also as cores for the compound connectives "as for" and "more crucially" respectively.

Orientative connectives of adjustment constitute the larger of the two categories both in terms of tokens and types. It has a relative frequency of 55% within orientatives and 6% within total connective tokens. Further, and more interestingly, it comprises

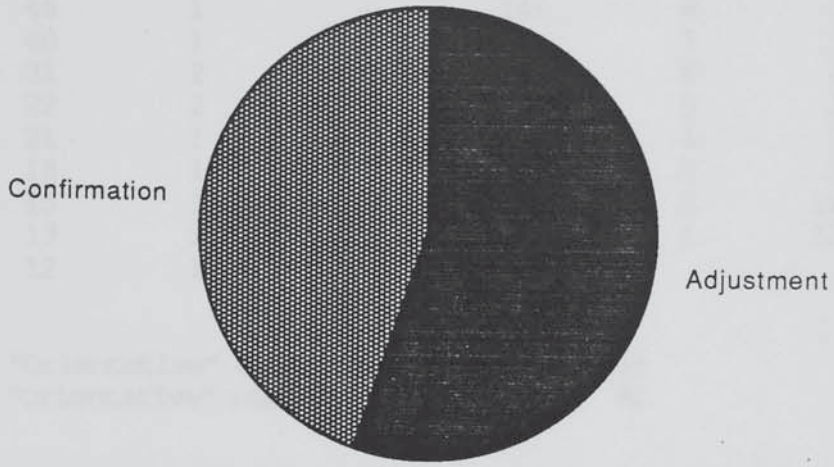
Category	Tokens	% Tokens	Types	% Types
Adjustment	559	55.29	103	79.84
Confirmation	452	44.71	26	20.16
Total	1011	100.00	129	100.00

Table 8.38 Distribution of categories of orientative connectives in the English corpus

80% of types of orientatives and 27% of categorised types. This is the highest share of types that any subcategory has within the corpus. The type token ratio in this category is 0.18.

The orientative category of confirmation consists of 452 tokens that constitute 45% of orientatives but comprises 26 types, i.e. only 20% of orientatives and 7% of categorised types. The type-token ratio in this category is 0.057, considerably lower than that of the orientative category of adjustment.

Fig. 8.11 Distribution of the Categories of Orientative Connectives in the English Corpus



8.6.2 Orientative Connectives in Arabic

8.6.2.1 Global Profile

1. Frequency

In the Arabic corpus the observed number of orientative connective types is 92 having a total frequency of 738. The type-token ratio is 0.125. The rank distribution (Table 8.39; see also the frequency lists in Appendices 87-88) consists of 21 ranks. The top rank is occupied by the connective "'ammā" with a frequency of 131 that is equivalent to 17.75% of total orientative tokens. The next rank is taken up by "kamā" (used for assigning authority, see Ch. 6), with a frequency of 88 that corresponds to 12% of orientative coverage. Some other frequent orientatives include "lā

Rank	Tokens	Types	Ranks	Tokens	Types
1:	131	1	12:	11	2
2:	88	1	13:	9	4
3:	46	1	14:	8	3
4:	40	1	15:	7	5
5:	31	2	16:	6	2
6:	22	2	17:	5	2
7:	21	1	18:	4	3
8:	18	1	19:	3	8
9:	15	1	20:	2	11
10:	13	2	21:	1	38
11:	12	1			

Number of "Orientative" connective tokens: 738
Number of "Orientative" connective types : 92

Table 8.39 Rank distribution of orientative connectives in the Arabic corpus

šakka" ("šakka" being the core element of this compound connective; it has a frequency of 46 which is equal to 6% of total orientative tokens), "fi<lan" (40 occurrences), "innamā" and "xāṣṣatan" (both occupying the same rank, with a frequency of 31 each) and "<alā al-'aqalli" and "bi-al-ṭab<i" (both filling the same rank and having a frequency of 22 each). Distributions of orientative connectives first within the corpus and then in relation to total connective tokens are displayed in Appendices (89-90) respectively.

The first 8 ranks in the distribution achieve collectively slightly more than 61% of orientative token coverage (standing for 450 tokens), though only around 11% of orientative types. The lowest rank is taken up by 38 hapaxes that correspond to 41% of types but achieve 5% of token coverage.

2. Repetitiveness

Indices of repetitiveness have been computed as follows:

a. Orientative connective occurrence rate is 23. This indicates that one orientative connective is encountered in 23 successive connective tokens.

b. Type occurrence rate in this category is 8, that is one new type appears in every 8 occurrences of orientatives.

c. General repeat rate is (19×10^{-4}). That is, the probability that two successive connectives - with or without a gap separating them - are any orientatives is 19 in 10,000.

d. Orientative system repeat rate is (12×10^{-5}). This indicates that the probability that two successive connectives are the same orientative type is 12 in 100,000.

e. The computation of gap distribution indicates that the average length of distances between two orientatives in the corpus is 346 words. The shortest distance has a length of 0 words (two orientatives follow each other immediately with no intervening words) and a frequency of 2. The longest gap has length of 3,961 words and a frequency of one. The most frequent distance occurs 6 times and represents 5 different distance lengths: 19, 38, 41, 91 and 216.

3. Growth

Results of the computation of global and local growth are given in Tables (8.40-41) and plotted in Appendices (91-92).

a. Global Growth: This measure computes the growth of orientatives within successive text intervals, each made up of 5,000

Tuldava's (token) richness index: 6.089089
Growth rate of "orientative" tokens: -6.047555
Tuldava's (type) richness index: 1.073594
Growth rate of "orientative" types: -4.028952

Actual Text Tokens	Actual Orien.Tokens	Expected Orien.Tokens	Actual Orien.Types	Expected Orien.Types
5000	15	12	9	10
10000	30	25	14	14
15000	37	38	17	18
20000	43	52	19	22
25000	60	65	26	25
30000	67	79	27	28
35000	84	93	32	31
40000	98	107	34	33
45000	117	121	38	36
50000	137	134	39	38
55000	151	148	43	40
60000	167	162	47	43
65000	186	177	50	45
70000	207	191	52	47
75000	217	205	57	49
80000	223	219	58	51
85000	236	233	60	53
90000	242	247	62	55
95000	256	262	64	57
100000	264	276	65	59
105000	279	290	66	60
110000	297	304	67	62
115000	303	319	67	64
120000	320	333	70	66
125000	339	347	70	67
130000	365	362	74	69
135000	384	376	77	71
140000	393	391	77	72
145000	407	405	78	74
150000	419	419	78	75
155000	433	434	79	77
160000	449	448	80	78
165000	472	463	80	80
170000	484	477	83	81
175000	505	492	84	83
180000	521	506	84	84
185000	528	521	84	86
190000	542	535	86	87
195000	556	550	87	89
200000	577	564	89	90
205000	593	579	89	92
210000	612	594	89	93
215000	621	608	89	94

220000	631	623	89	96
225000	648	637	90	97
230000	670	652	90	99
235000	677	667	90	100
240000	688	681	90	101
245000	704	696	90	103
250000	719	711	90	104
255000	735	725	93	105
256450	738	729	94	106

Extrapolated Text Tokens	Expected Orien. Tokens	Expected Orien. Types
500000	1451	161
750000	2202	209
1000000	2959	251

Table 8.40 Global growth of orientative connectives in the Arabic corpus

words. The first interval includes 15 orientative tokens, which contain 9 types. The expected numbers, using Tuldava's formula, are 12 tokens and 10 types. The next interval includes another 15 tokens, raising the number to 30, and 4 new types, raising the number to 14. Growth of tokens fluctuates between 26, the highest level, at interval 26, to 6, the lowest level, at intervals 4, 16, 18 and 23 (excluding the final interval which is shorter than the rest). Growth of types is steady, with the 50th type percentile occurring early in the corpus, at interval 12 (60,000 words). Growth then drops slightly and signs of near saturation appear mid-way in the corpus. For instance, the 75th type percentile occurs within interval 24 (120,000 words). An extrapolated corpus of 500,000 tokens is expected to contain 1,451 orientative tokens and 161 types. An extrapolated corpus of a million words is expected to include 2,959 orientative tokens and 251 types.

Tuldava's richness index:	4.066002
"Orientative" Connective growth rate:	-3.096516

Actual Connective Tokens	Actual "Orientative" Connective Tokens	Expected "Orientative" Connective Tokens
500	20	16
1000	37	36
1500	50	55
2000	66	76
2500	88	96
3000	112	117
3500	138	139
4000	162	160
4500	189	181
5000	212	203
5500	222	225
6000	242	247
6500	263	269
7000	294	291
7500	308	313
8000	339	335
8500	378	357
9000	393	380
9500	413	402
10000	429	425
10500	456	447
11000	474	470
11500	501	492
12000	521	515
12500	537	538
13000	558	561
13500	585	584
14000	612	607
14500	630	630
15000	666	653
15500	676	676
16000	698	699
16500	718	722
16995	738	745

Extrapolated Connective Tokens	Expected "Orientative" Connective Tokens
20000	884
30000	1356
50000	2320
100000	4798

Table 8.41 Local growth of orientative connectives in the Arabic corpus

b. Local Growth: This measure monitors the growth of orientative tokens within intervals each containing 500 connective tokens. The first interval includes 20 orientatives (the expected figure is 16). The next interval contains 17 more orientatives, raising the number to 37. Growth fluctuates between 39, the highest level, at interval 17, to 10 at intervals 11 and 31. An extrapolation of a corpus containing 20,000 connectives is expected to include 884 orientative tokens. Extrapolation of a larger corpus that contains 100,000 connectives is expected to have 4,798 orientatives.

8.6.2.2 Categories of Orientative Connectives in Arabic

The distribution of the categories of orientative connectives is given in Table (8.42) and displayed in Figure (8.12). More detailed results of computation are given in the frequency lists in Appendices (93A-B). The total number of observed types in both categories of orientative connectives is 93, an increase of one type over global orientatives. This is because the core "ḥāli" stands for the two connectives: "fī hāḏihi al-ḥāli" and "bi-ṭabī<ati al-ḥāli".

Category	Token	% Token	Type	% Type
Adjustment	465	63	66	71
Confirmation	273	37	27	29
Total	738	100	93	100

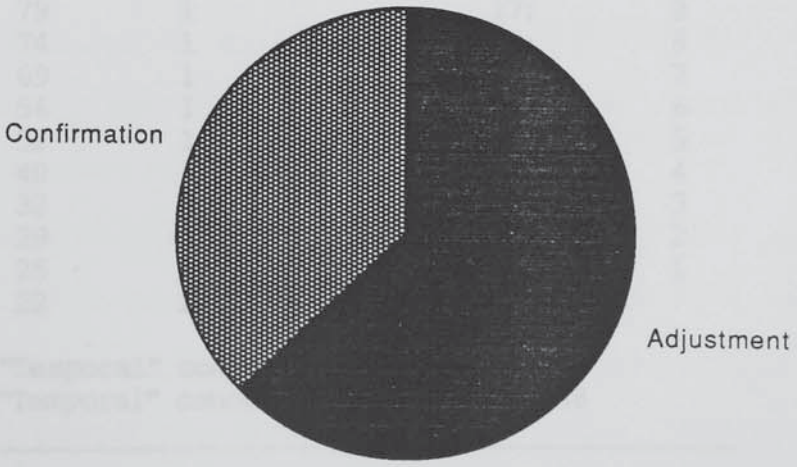
Table 8.42 Distribution of the categories of orientative connectives in the Arabic corpus

The orientative connectives of adjustment form the larger of the two orientative categories in terms of tokens and types. They

comprise 64 types, representing 71% of orientative types and 22% of total connective types at this level of categorisation. The category has a frequency of 462, corresponding to 63% of total orientative coverage, and approximately 3% of total connectives in the corpus. The most frequent connective in this category is "'ammā", which achieves a frequency of 131, that is 28% of orientative/adjustment connectives (see App. 86A).

The orientative category of confirmation has a frequency of 273 representing 37% of total orientatives. In terms of types, it comprises 27 types that constitute 29% of total orientative types at this level of categorisation. The most frequent connective in this category is "lā šakka", which as a frequency of occurrence of 46 representing 17% of tokens in this category (see App. 86B for more details).

Fig. 8.12 Distribution of the Categories of Orientative Connectives in the Arabic Corpus



8.7 Temporal Connectives

8.7.1 Temporal Connectives in English

8.7.1.1 Global Profile

1. Frequency

Temporal connectives in the English corpus comprise 48 types that have a total frequency of 1,284. The rank distribution (Table 8.43, see also the frequency lists in Appendices 94-95) consists of 25 ranks. The top rank is taken up by the connective "when" with a frequency of 414 that is equivalent to 32% of temporal tokens and 4% of total connectives. The next rank is filled by the connective "then" with a frequency of 127 that corresponds to 10% of temporals. The next three ranks are occupied by "as" (86 occurrences, 7% of temporals), temporal "and" (79 occurrences, 6% of temporals), and "before" (74 occurrences, 6% of temporals).

Rank	Tokens	Types	Rank	Tokens	Types
1:	414	1	14:	17	1
2:	127	1	15:	16	1
3:	86	1	16:	10	1
4:	79	1	17:	9	2
5:	74	1	18:	8	1
6:	69	1	19:	7	2
7:	56	1	20:	6	5
8:	52	1	21:	5	2
9:	40	1	22:	4	1
10:	32	1	23:	3	3
11:	29	1	24:	2	6
12:	25	1	25:	1	9
13:	22	2			

Number of "Temporal" connective tokens: 1284
Number of "Temporal" connective types : 48

Table 8.43 Rank distribution of temporal connectives in the English corpus

These five ranks achieve more than 60% of temporal token coverage. The lowest rank is occupied by 9 hapaxes that represent 18.75% of types, though only 0.7% of tokens. The type-token ratio in this category is 0.037. The frequency distribution of temporals within the corpus is given in Appendix (96) and their distribution within total connective tokens is displayed in Appendix (97).

2. Repetitiveness

Indices of repetitiveness of temporal connectives in the English corpus have been computed as follows:

a. Occurrence rate of temporal connectives is 7.5. This indicates the average number of successive connectives we have in a text chunk in order to encounter a temporal connective.

b. Type occurrence rate within this category is 3. In other words, one new type is encountered in every 3 successive temporal connectives in the corpus.

c. General repeat rate is 0.018. This refers to the probability that two successive connective tokens - with or without a gap separating them - are any temporals.

d. Temporal system repeat rate is (24×10^{-4}). This refers to the probability that two successive connectives in the corpus turn out to be the same temporal type.

e. Results of the gap distribution show that the average gap length is 199 words. The shortest gap length is 0 words (two temporal tokens follow each other immediately, i.e. without intervening words) and has a frequency of 13. The longest gap length is 1,548 words. The most frequent gaps are 8 and 16 words, each with 14 occurrences.

3. Growth

Results of the calculation of growth are recorded in Tables (8.44-45) and plotted in Appendices (98-99).

a. Global Growth: Growth is observed in successive intervals of text segment, each equal to 5000 words. The first interval contains 24 temporal tokens and 13 types. The expected numbers (arrived at by using Tuldava's formula) are 26 tokens and 16 types. In the next interval tokens increase to 54 (a growth of 30) and types to 18 (a growth of 5). Growth of tokens fluctuates between 39, the highest level of growth, achieved at interval 28, to 11, growth lowest level, achieved at interval 8. Types grow fast and therefore get saturated at an early interval. For instance the 50th type percentile occurs very early in the corpus; it is achieved within interval 28 (140,000 words). After that, growth slumps and signs of saturation are evident.

Tuldava's (token) richness index:	5.002819
Growth rate of "temporal" tokens:	-5.012152
Tuldava's (type) richness index:	0.615797
Growth rate of "temporal" types:	14.005726

Actual Text Tokens	Actual Temp.Tokens	Expected Temp.Tokens	Actual Temp.Types	Expected Temp.Types
5000	24	26	13	16
10000	54	53	18	19
15000	74	79	22	22
20000	107	105	26	24
25000	134	131	29	25
30000	163	157	30	27
35000	196	184	30	28
40000	207	210	31	29
45000	240	236	33	30
50000	268	262	33	31
55000	296	288	35	32

60000	324	313	35	33
65000	348	339	35	34
70000	376	365	35	34
75000	400	391	37	35
80000	431	417	38	36
85000	460	443	38	36
90000	479	469	38	37
95000	515	495	38	37
100000	534	521	39	38
105000	552	546	40	39
110000	580	572	41	39
115000	602	598	41	40
120000	626	624	42	40
125000	643	650	42	41
130000	673	675	42	41
135000	698	701	42	42
140000	737	727	43	42
145000	757	753	43	42
150000	787	778	43	43
155000	814	804	43	43
160000	840	830	43	44
165000	870	856	44	44
170000	893	881	45	44
175000	908	907	45	45
180000	929	933	46	45
185000	955	959	46	46
190000	981	984	46	46
195000	998	1010	46	46
200000	1022	1036	46	47
205000	1045	1061	46	47
210000	1061	1087	46	47
215000	1078	1113	46	48
220000	1107	1138	46	48
225000	1133	1164	46	48
230000	1161	1190	46	48
235000	1182	1216	46	49
240000	1213	1241	47	49
245000	1241	1267	48	49
250000	1260	1293	48	50
255000	1277	1318	48	50
256560	1284	1326	48	50

Extrapolated Text Tokens	Expected Temp. Tokens	Expected Temp. Types
500000	2572	61
750000	3847	68
1000000	5119	74

Table 8.44 Global growth of temporal connectives in the English corpus

Tuldava's richness index:	1.070289
"Temporal" Connective growth rate:	-1.084159

Actual Connective Tokens	Actual "Temporal" Connective Tokens	Expected "Temporal" Connective Tokens
500	70	70
1000	142	139
1500	205	207
2000	276	275
2500	347	342
3000	409	409
3500	477	476
4000	546	542
4500	604	609
5000	669	675
5500	746	742
6000	814	808
6500	880	874
7000	938	940
7500	999	1006
8000	1069	1072
8500	1135	1138
9000	1204	1204
9500	1273	1270
9596	1284	1282

Extrapolated Connective Tokens	Expected "Temporal" Connective Tokens
10000	1336
15000	1991
20000	2642
30000	3940
50000	6518
100000	12912

Table 8.45 Local growth of temporal connectives in the English Corpus

On the basis of the growth of temporal types and tokens, an extrapolated corpus of 500,000 words is expected to include 2,572 temporal tokens and 61 types. Extrapolation of a bigger corpus is expected to contain 5,119 tokens and 74 types.

2. Local Growth: Growth of temporal connective tokens is calculated within successive text intervals each containing 500 connective tokens. The first interval includes 70 temporals, which is the same as the expected number. The next interval contains 72 further tokens bringing the total to 142. Growth varies between 77, the highest level of growth, achieved at interval 11, to 58, growth lowest level, at intervals 9 and 14. An extrapolated text corpus containing 10,000 connective tokens is expected to include 1,336 temporal tokens. Extrapolation of a corpus containing 100,000 connective tokens is expected to contain 12,912 temporals.

8.7.1.2 Categories of Temporal Connectives in English

The distribution of the various categories of temporal connectives is summarised in Table (8.46) and plotted in Figure (8.13). Further detailed frequency lists of each temporal category are given in Appendices (100A-E). At this level of categorisation,

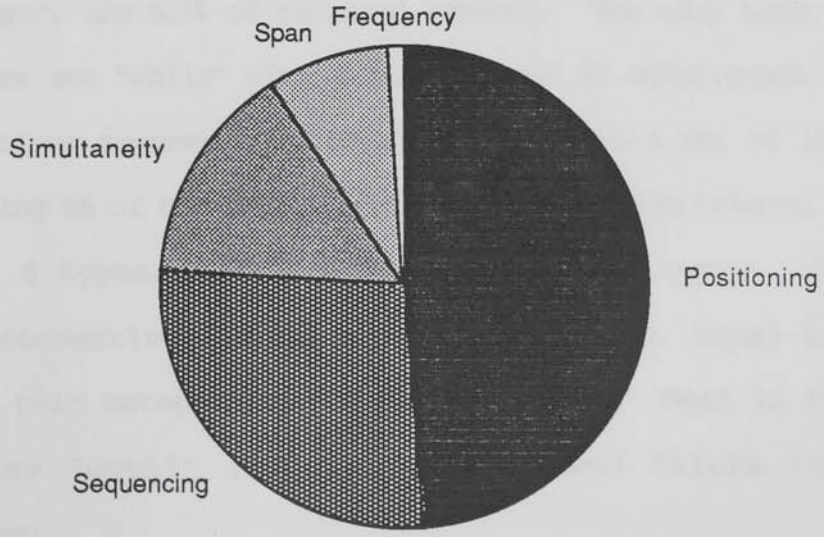
Category	Tokens	% Tokens	Type	% Type
Positioning	626	48.75	22	37.29
Sequence	356	27.73	20	33.90
Simultaneity	187	14.56	9	15.25
Span	104	8.10	6	10.17
Frequency	11	0.86	2	3.39
Total	1284	100.00	59	100.00

Table 8.46 Distribution of the categories of temporal connectives in the English corpus

the number of types is 59, an increase of 11 types over global temporal types. This increase is due to the following: the connectives "as" and "when" expressing temporal positioning and simultaneity, "and" expressing sequence and simultaneity, the core "time" entering into these compound connectives "all this time", "this time", "at one time" and "during a time", the connective "then" and "later" expressing positioning and sequence, the connective "first" acting as a core in "first of all", the connective "then" acting as a core in "by then", the connective "since" as a core in "since then", and the connective "after" as a core in the compound connective "after this".

Connectives expressing temporal positioning consist of 626 tokens, which represents approximately half temporal tokens, and comprise 22 types, of which 5 are hapaxes. The most frequent

Fig. 8.13 Distribution of the Categories of Temporal Connectives in the English Corpus



connective is "when". It has a frequency of 413 which stands for 66% of tokens of temporal positioning, 32% of total temporal tokens and 4.3% of global connective tokens. The next most frequent connective is "then" (62 occurrences, equal to 11% of temporal positioning tokens and 4.8% of total temporals).

Connectives expressing sequence consist of 356 tokens that correspond to 28% of temporals and 3.7% of total connective tokens, and comprise 20 types at this level of categorisation, of which 2 are hapaxes. The most frequent connective is sequential "and" (76 occurrences, standing for 21% of sequential connectives and about 6% of temporal tokens). The next most frequent connectives are "before" (74 occurrences), "then" (63 occurrences) and "after" (29 occurrences).

Connectives expressing temporal simultaneity make up 187 tokens that represent approximately 15% of temporals and about 2% of total connective tokens. The set comprises 9 types (equal to 15.25% of temporal types, of which 2 are hapaxes. The most frequent connective is "as" (71 occurrences equal to 37% of connectives in this category and 5.5% of temporal tokens). The next most frequent connectives are "while" with a frequency of 52 occurrences.

Connectives expressing temporal span make up a set of 104 tokens representing 8% of temporals and 1% of connective tokens. The set comprises 6 types, equal to 10% of temporal types. The most frequent connective is "since" (42 occurrences, equal to 40% of tokens in this category and 3% of temporals). Next in frequency order come "until" (40 occurrences) and "since then" (14 occurrences).

The last temporal category in our scheme is that of frequency.

This is a small category which consists of 11 tokens and 2 types: "whenever" (9 occurrences) and the core "time" in "every time" (2 occurrences).

8.7.2 Temporal Connectives in Arabic

8.7.2.1 Global Profile

1. Frequency

Computation of temporal connectives in the Arabic corpus identifies 62 types that have a total frequency of 1,494. The rank distribution (Table 8.47; see also frequency lists in Appendices 101-102) is composed of 25 ranks. The top rank is taken up by the connective "wa", with a frequency of 437 that is equivalent to 29% of total temporal tokens. The next 4 ranks are occupied respectively by: "ba<da" (157 occurrences, equivalent to 10.5% of temporal tokens), "<indamā" (146 occurrences, 10% of temporal

Rank	Tokens	Types	Rank	Tokens	Types
1:	436	1	14:	18	2
2:	157	1	15:	15	1
3:	146	1	16:	12	1
4:	95	1	17:	11	1
5:	80	1	18:	10	1
6:	66	1	19:	9	4
7:	65	1	20:	7	2
8:	64	1	21:	5	4
9:	52	1	22:	4	3
10:	27	1	23:	3	7
11:	23	1	24:	2	3
12:	21	1	25:	1	19
13:	20	2			

Number of "Temporal" connective tokens: 1484
Number of "Temporal" connective types : 62

Table 8.47 Rank distribution of temporal connectives in the Arabic corpus

tokens), "ṯumma" (95 occurrences, 6.4% of temporal tokens) and "fa" (80 occurrences, 5.4% of temporal tokens). These five connectives achieve collectively more than 61% of temporal token coverage. The lowest rank is occupied by 19 hapaxes that represents about 31% of total temporal types, though only 1.3% of temporal tokens. The type-token ratio in this category is 0.041. Distributions of temporals first in relation to text tokens and then to total connective tokens are displayed in Appendices (103-104).

2. Repetitiveness

Indices of repetitiveness of temporal connectives in the Arabic corpus are computed as follows:

a. Occurrence rate of temporal connectives is 11. This means that one temporal token is expected in an average number of 11 successive connective tokens.

b. Type occurrence rate within this category is 24. This refers to the average number of successive temporal tokens we must have before we encounter a new type.

c. General repeat rate of Arabic temporal connectives is 0.008. This refers to the probability that two successive tokens - with or without a gap separating them - turn out to be temporals.

d. Temporal system repeat rate is (94×10^{-5}). This index measures the probability that two successive connective tokens turn out to be the same temporal type.

e. The gap distribution indicates that the average gap length is 171 words. The shortest gap length is 0 and has 23 occurrences. The longest gap length is 2,541 words and has one occurrence. The

most frequent gap length is 8 words and has a frequency of 32.

3. Growth

Results of the calculation of global and local growth of temporal connectives in the Arabic corpus are given in Table (8.48-49) and plotted in Appendices (105-106).

a. Global Growth: Growth of temporal tokens and types are monitored in text intervals, each 5,000 words of length. The first interval contains 37 temporal tokens, 2 tokens lower than the expected number. The number of types observed at this interval is 11, equal to the expected number. The next interval includes 40 more tokens, raising the number to 77, which is 4 tokens more than the expected number. Growth of types drops at this interval to 3.

Tuldava's (token) richness index:	3.018229
Growth rate of "temporal" tokens:	-3.095649
Tuldava's (type) richness index:	0.993448
Growth rate of "temporal" types:	-6.036053

Actual Text Tokens	Actual Temp.Tokens	Expected Temp.Tokens	Actual Temp.Types	Expected Temp.Types
5000	37	39	11	11
10000	77	73	14	15
15000	113	105	19	18
20000	155	136	22	20
25000	179	167	24	22
30000	206	197	24	24
35000	232	227	27	26
40000	256	256	30	27
45000	298	285	32	28
50000	316	313	33	30
55000	343	342	34	31
60000	369	370	35	32
65000	391	398	35	33
70000	420	426	35	34
75000	447	453	35	35
80000	479	481	35	36

85000	505	508	38	37
90000	515	535	38	38
95000	523	562	39	39
100000	550	589	39	40
105000	561	616	39	41
110000	588	643	42	42
115000	602	670	42	43
120000	632	696	43	43
125000	648	723	43	44
130000	670	749	43	45
135000	710	775	45	46
140000	734	802	45	46
145000	774	828	45	47
150000	795	854	45	48
155000	832	880	46	49
160000	865	906	46	49
165000	915	932	48	50
170000	945	958	49	51
175000	991	983	49	51
180000	1029	1009	51	52
185000	1068	1035	51	52
190000	1103	1060	52	53
195000	1138	1086	54	54
200000	1177	1111	55	54
205000	1220	1137	58	55
210000	1234	1162	59	55
215000	1270	1188	60	56
220000	1286	1213	60	57
225000	1310	1238	60	57
230000	1347	1263	61	58
235000	1375	1289	61	58
240000	1394	1314	61	59
245000	1416	1339	61	59
250000	1443	1364	62	60
255000	1471	1389	62	60
256450	1494	1396	62	60

Extrapolated Text Tokens	Expected Temp. Tokens	Expected Temp. Types
500000	2581	81
750000	3751	97
1000000	4894	110

Table 8.48 Global growth of temporal connectives in the Arabic corpus

Tuldava's richness index:	1.020022
"Temporal" Connective growth rate:	-1.077041

Actual Connective Tokens	Actual "Temporal" Connective Tokens	Expected "Temporal" Connective Tokens
500	61	57
1000	108	106
1500	169	153
2000	202	199
2500	239	243
3000	291	287
3500	321	330
4000	364	373
4500	400	415
5000	430	456
5500	478	498
6000	515	539
6500	547	580
7000	578	620
7500	618	661
8000	648	701
8500	697	741
9000	737	780
9500	782	820
10000	830	859
10500	871	898
11000	935	938
11500	984	976
12000	1032	1015
12500	1084	1054
13000	1140	1093
13500	1191	1131
14000	1234	1169
14500	1282	1208
15000	1316	1246
15500	1373	1284
16000	1406	1322
16500	1443	1359
16995	1494	1397

Extrapolated Connective Tokens	Expected "Temporal" Connective Tokens
20000	1622
30000	2355
50000	3772
100000	7160

Table 8.49 Local growth of temporal connectives in the Arabic corpus

In general, growth fluctuates within the successive intervals. Growth of tokens varies from 50, growth highest level, achieved at interval 33 to 8, growth lowest level, at interval 19. Growth of tokens drops after the first interval, and remains slow but inconsistent. For instance the 50th type percentile occurs with interval 9 (45,000 words) but the 75th type percentile occurs at interval 31 (155,000 words). The 90th type percentile occurs within interval 40 (200,000 words).

Extrapolation on the basis of growth gives the following results. An extrapolated corpus of 500,000 words is expected to include 2,581 temporal tokens and 81 types. An extrapolated corpus of a million words is estimated to comprise 4,894 temporal tokens and 110 types.

b. Local Growth: Growth of temporal tokens is monitored within text intervals each comprising 500 connective tokens. The first interval contains 61 connective tokens (the expected number is 57). Growth after this fluctuates between 64, growth highest level, at interval 22, to 30, growth lowest level, at interval 10. An extrapolated text corpus containing 20,000 connective tokens is expected to comprise 1,622 temporal tokens. An extrapolated corpus that contains 100,000 connectives is expected to include 7,160 temporals.

8.7.2.2 Categories of Temporal Connectives in Arabic

Distribution of categories of temporal connectives in the Arabic corpus is summarised in Table (8.50) and plotted in Figure (8.14). Results of further calculations for each temporal category are recorded in Appendices (107A-E).

At this level of categorisation the number of temporal types observed in the corpus is 73, an increase of 11 types over global temporal types. This is due to the following factors: a) multi-functionality of the temporal connectives "wa" (sequence and circumstance), "ḥattā" (positioning and span) and "min qablu" and "fī al-bidāyati" (both expressing positioning and sequence), "'iḏ" and "fī waqtin" (both expressing positioning and simultaneity);

Category	Token	% Token	Type	% Type
Positioning	400	26.77	37	50.68
Sequence	780	52.21	18	24.66
Simultaneity	62	4.15	7	9.59
Span	110	7.36	6	8.22
Frequency	17	1.14	4	5.48
Circumstance	125	8.37	1	1.37
Total	1494	100.00	73	100.00

Table 8.50 Distribution of categories of temporal connectives in the Arabic corpus

b) the multi-use of the following cores, forming simple and multi-word connectives: "ba<da", "waqti" (in "fī al-waqti nafsi-hi" and "fī hāḏā al-waqti"), "kullamā" (simple connective and in the correlate "kullamā ... kullamā"), "munḏu" (both as simple connective and in the compound "munḏu ḏālika al-waqti").

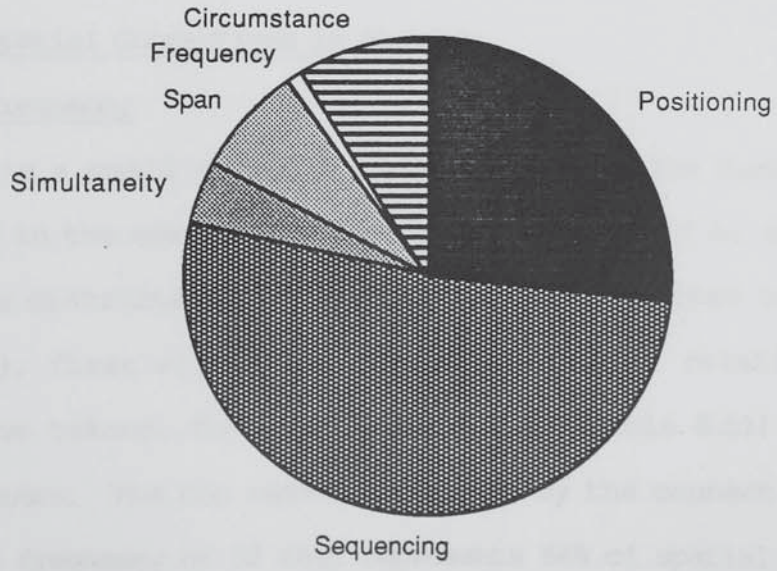
Connectives of temporal positioning consist of 400 tokens (equivalent to 27% of temporal tokens and 2.3% of total connectives) and comprise 37 types, of which 13 are hapaxes. The most frequent connective is "<indamā" (having a frequency of 146, equal to 37% of tokens in this category, 9.84% of temporals). Next in the order of frequency come "ḥīna" (52 occurrences) and "ḥīnamā" (21 occurrences).

Connectives of sequence form the biggest temporal category in

terms of tokens: 780 tokens that stand for 52% of temporals and 4.6% of total connectives. The category comprises a set of 18 types, of which 6 are hapaxes. The most frequent connective in the set is sequential "wa", which has a frequency of 311 representing 40% of sequential tokens, 21% of temporals and 1.8% of total connectives. The next five connectives in order of descending frequency are "ba<da" (156 occurrences, equivalent to 20% of sequentials and 10% of temporals), "ṯumma" (95 occurrences) and "ʾaxīran" (18 occurrences).

Connectives of simultaneity consist of 61 tokens that form 4% of temporal connectives and 0.36% of total connective tokens. The set comprises 7 types, of which 3 are hapaxes. The most frequent connective is "fī al-waqti nafsi-hi", which has a frequency of 45, representing 74% of tokens of this category and 3% of temporals.

Fig. 8.14 Distribution of the Categories of Temporal Connectives in the Arabic Corpus



The next set of connectives are those expressing temporal span. They consist of 110 tokens that form 7% of total temporals and 0.65% of total connective tokens. The set comprises 6 types, of which 2 are hapaxes. The most frequent connective in the set is "ḥattā", which has a frequency of 55, representing 50% of tokens in the set and 3.7% of temporals. The next three most frequent connectives are "'ilā 'an" (27 occurrences) and "mundu" (22 occurrences).

The smallest temporal category in terms of tokens is that of frequency. It comprises 17 tokens, representing 1% of temporal tokens, and has 4 types, of which 2 are hapaxes. These types are "kullamā" (13 occurrences), "kullamā ... kullamā" (as a correlate, 2 occurrences), "mā dāma" and "mā 'an" (1 occurrence each).

The last temporal category is that of circumstance. This category involves one type: circumstantial "wa". It has a frequency of 124 which constitutes 8.35% of temporal tokens and 0.73% of total connective tokens.

8.8 Spatial Connectives

8.8.1 Spatial Connectives in English

1. Frequency

This is a small category of connectives. The number of types observed in the corpus is 6 having a frequency of 50 (see Appendix 108). The distribution of the category is exhibited in Appendices (109-110), first within the corpus and then in relation to total connective tokens. The rank distribution (Table 8.51) consists of only 4 ranks. The top rank is occupied by the connective "where", having a frequency of 32 that represents 64% of spatial tokens. The

next two ranks are taken up by "elsewhere" (8 occurrences, 16% of spatial tokens) and "whenever" (7 occurrences, 14% of spatial tokens). The lowest rank is occupied by 3 hapaxes: "at this point", "therein" and "whereby". The type-token ratio of this category is 0.12.

Rank	Tokens	Types
1:	32	1
2:	8	1
3:	7	1
4:	1	3

Number of "Spatial" connective tokens: 50
Number of "Spatial" connective types : 6

Table 8.51 Rank distribution of spatial connectives in the English corpus

2. Repetitiveness

Indices of repetitiveness of spatial connectives are computed as follows:

a. Occurrence rate of spatial connectives is 192. This low size of repetitiveness is due to the small number of tokens in this category.

b. Type occurrence rate in this category is 8. That is a new type appears after 8 successive spatial tokens.

c. General repeat rate is (27×10^{-6}). This refers to the probability that two successive connectives - with or without a gap - are any spatial connectives.

d. Spatial system repeat rate is (124×10^{-7}). This is the probability that two successive connective tokens in the corpus are the same spatial connective.

e. Gap distribution shows that due to the limited number of connective tokens, the average distance length is very large: 5,026 words. Each distance length has a frequency of one. The shortest gap length consists of 5 words and the largest comprises 24,411 words.

3. Growth

Results of calculation of global and local growth of spatial connectives in the English corpus are given in Tables (8.52-53) and plotted in Appendices (111-112).

a. Global Growth: This measure traces growth of spatial connectives within intervals of 5,000 words each. The first interval does not contain any spatial connectives. The second interval contains 2 tokens, both different types. Growth is slow and inconsistent: some intervals contain up to 3 connectives each, others do not include any. An extrapolated corpus of 500,000 words is expected to have 96 spatial tokens (an increase of 46 tokens over

---

Tuldava's (token) richness index: 6.048227
Growth rate of "spatial" tokens: -7.026629
Tuldava's (type) richness index: 0.977065
Growth rate of "spatial" types: -18.000908

Actual Text Tokens	Actual Spat.Tokens	Expected Spat.Tokens	Actual Spat.Types	Expected Spat.Types
5000	0	0	0	0
10000	2	2	2	3
15000	4	3	4	3
20000	5	4	4	3
25000	6	6	5	4
30000	6	7	5	4
35000	9	8	5	4
40000	9	9	5	4
45000	11	10	5	4

50000	13	11	5	4
55000	15	12	5	4
60000	15	13	5	4
65000	16	14	5	4
70000	18	15	5	4
75000	18	16	5	4
80000	20	17	5	5
85000	20	18	5	5
90000	20	19	5	5
95000	20	20	5	5
100000	22	21	5	5
105000	22	22	5	5
110000	23	23	5	5
115000	24	24	5	5
120000	26	25	5	5
125000	27	26	5	5
130000	27	27	5	5
135000	27	28	5	5
140000	28	29	5	5
145000	29	30	6	5
150000	29	31	6	5
155000	30	32	6	5
160000	34	33	6	5
165000	35	34	6	5
170000	35	35	6	5
175000	36	36	6	5
180000	37	37	6	5
185000	39	38	6	5
190000	39	39	6	5
195000	42	40	6	5
200000	42	41	6	5
205000	42	42	6	6
210000	42	42	6	6
215000	42	43	6	6
220000	44	44	6	6
225000	44	45	6	6
230000	45	46	6	6
235000	47	47	6	6
240000	47	48	6	6
245000	48	49	6	6
250000	49	50	6	6
255000	50	51	6	6
256560	50	51	6	6

Extrapolated Text Tokens	Expected Spat. Tokens	Expected Spat. Types
500000	96	7
750000	140	7
1000000	183	8

Table 8.52 Global Growth of spatial connectives in the English corpus

Tuldava's richness index:	3.045886
"Spatial" Connective growth rate:	-4.027037

Actual Connective Tokens	Actual "Spatial" Connective Tokens	Expected "Spatial" Connective Tokens
500	4	3
1000	6	6
1500	9	9
2000	14	12
2500	16	15
3000	18	17
3500	20	20
4000	22	22
4500	24	25
5000	27	27
5500	29	30
6000	31	32
6500	35	34
7000	38	37
7500	42	39
8000	42	41
8500	44	44
9000	47	46
9500	50	48
. 9596	50	49

Extrapolated Connective Tokens	Expected "Spatial" Connective Tokens
10000	51
15000	73
20000	95
30000	137
50000	217
100000	407

Table 8.53 Local growth of spatial connectives in the English corpus

the observed number in the present corpus) and 7 types (only one new type). An extrapolated corpus of a million words is expected to contain 183 spatial tokens and 8 types.

b. Local Growth: Growth of spatial connectives is also monitored in text intervals each containing 500 connectives. The first interval contains 4 spatial tokens. Growth varies across the intervals: one interval contains 5 (interval 4) while another (interval 16) does not have any. An extrapolated corpus containing 10,000 connective tokens is expected to include 51 spatial tokens. An extrapolated corpus containing 100,000 connectives is expected to contain 407.

8.8.2 Spatial Connectives in Arabic

1. Frequency

The number of spatial connective types observed in the corpus is 3, having a total frequency of 40. These figures make this category the smallest in the corpus (see Appendix 113). The distribution of the category is displayed in Appendices (114-115), first within the entire corpus, then in relation to total connective tokens. The rank distribution (Table 8.54) is composed of 3 ranks, each filled

Rank	Tokens	Types
1:	37	1
2:	2	1
3:	1	1

Number of "Spatial" connective tokens: 40
Number of "Spatial" connective types : 3

Table 8.54 Rank distribution of spatial connectives in the Arabic corpus

by only one type. The top rank is occupied by the connective "ḥaytu". It has a frequency of 36, which is equivalent to 92% of total spatial tokens. The next rank is reserved for the connective "haytuma" with a frequency of 2 and the lowest rank has one hapax "fī hāḏā al-majāli" (the core element is "majali"). The type-token ratio in this category is 0.075.

2. Repetitiveness

Some indices of repetitiveness have been computed as follows:

a. Occurrence rate of spatial connectives is 425. This refers to the average number of connective tokens we may have before encountering a spatial connective. This low rate of repetitiveness is due to the small size of the category.

b. Type occurrence rate is 13. That is, one type appears within an average number of 13 successive spatial tokens.

c. General repeat rate is (55×10^{-7}). This means that the probability that two successive connectives - with or without a gap - are any spatials is 55 in 10 million.

d. Spatial system repeat rate is (48×10^{-7}). This refers to the probability that two successive connective tokens are the same spatial type.

e. Calculation of gap distribution indicates that the average distance length between 2 spatial connectives is very large: 6,253. The shortest distance consists of 8 words while the longest is 44,799 words.

3. Growth

Results of calculation of growth are given in Tables (8.55-56) and plotted in Appendices (116-117).

Tuldava's (token) richness index: 25.008271
Growth rate of "spatial" tokens: -18.001503
Tuldava's (type) richness index: 1.097966
Growth rate of "spatial" types: -6.079714

Actual Text Tokens	Actual Spat.Tokens	Expected Spat.Tokens	Actual Spat.Types	Expected Spat.Types
5000	1	0	1	0
10000	1	0	1	0
15000	1	0	1	0
20000	1	1	1	0
25000	1	1	1	0
30000	1	1	1	0
35000	1	2	1	1
40000	1	2	1	1
45000	1	3	1	1
50000	2	4	1	1
55000	2	4	1	1
60000	3	5	1	1
65000	3	5	1	1
70000	3	6	1	1
75000	5	6	1	1
80000	5	7	1	1
85000	6	8	1	1
90000	9	8	1	1
95000	9	9	1	1
100000	11	10	1	1
105000	12	10	1	1
110000	12	11	1	1
115000	14	12	1	1
120000	15	13	1	1
125000	16	13	1	1
130000	18	14	1	1
135000	19	15	1	1
140000	19	15	1	1
145000	19	16	1	1
150000	20	17	1	1
155000	23	18	1	1
160000	23	19	1	1
165000	23	19	1	1
170000	27	20	1	1
175000	27	21	1	1
180000	27	22	1	1
185000	30	23	2	1
190000	32	23	3	1
195000	32	24	3	1
200000	32	25	3	1
205000	32	26	3	1
210000	33	27	3	1

215000	33	28	3	1
220000	33	28	3	1
225000	33	29	3	1
230000	34	30	3	1
235000	34	31	3	1
240000	36	32	3	1
245000	36	33	3	1
250000	38	34	3	1
255000	40	35	3	1
256450	40	35	3	1

Extrapolated Text Tokens	Expected Spat. Tokens	Expected Spat. Types
500000	83	2
750000	140	2
1000000	202	2

Table 8.55 Global growth of spatial connectives in the Arabic corpus

a. Global Growth: Growth of spatial connectives are monitored in text intervals of 5,000 words each. The first interval contains one spatial token. No other token is encountered until interval 10, where another token appears. Growth is therefore both slow and inconsistent. An extrapolated corpus of 500,000 words is expected to contain 83 spatial tokens. An extrapolated corpus of a million words is expected to include 202. In both cases the expected number of types is 2, lower than the observed number of types in the corpus. This discrepancy is related to chance fluctuation.

b. Local Growth: This measure examines the growth of spatial connectives in text intervals, each containing 500 tokens. The first interval contains one spatial connective. Growth ceases for the next 5 intervals until another token appears in interval 7. An

Tuldava's richness index:	20.008836
"Spatial" Connective growth rate:	-13.005716

Actual Connective Tokens	Actual "Spatial" Connective Tokens	Expected "Spatial" Connective Tokens
500	1	0
1000	1	0
1500	1	1
2000	1	1
2500	1	2
3000	1	3
3500	2	4
4000	3	5
4500	3	5
5000	4	6
5500	5	7
6000	9	8
6500	11	9
7000	12	11
7500	14	12
8000	16	13
8500	19	14
9000	19	15
9500	19	16
10000	23	18
10500	23	19
11000	26	20
11500	27	21
12000	27	23
12500	32	24
13000	32	25
13500	32	27
14000	33	28
14500	33	30
15000	33	31
15500	34	32
16000	36	34
16500	38	35
16995	40	37

Extrapolated Connective Tokens	Expected "Spatial" Connective Tokens
20000	46
30000	78
50000	153
100000	370

Table 8.56 Calculation of "local" growth of spatial connectives in the Arabic corpus

extrapolated text corpus containing 20,000 connective tokens is expected to give 46 spatial tokens. An extrapolated text corpus containing 100,000 connective tokens is expected to include 370 spatials.

8.9 Causal Connectives

8.9.1 Causal Connectives in English

8.9.1.1 Global Profile

1. Frequency

Causal connectives observed in the corpus comprise 45 types that have a total frequency of 1,888. Details are given in the frequency lists in Appendices (118-119). The distribution is visually represented in the figures in Appendices (120-121). The rank distribution (Table 8.57) consists of 23 ranks. The first 14 ranks, having a total frequency of 1,307, constitute slightly less than 70% of total causal tokens.

Rank	Tokens	Types	Rank	Tokens	Types
1:	660	1	13:	13	1
2:	247	1	14:	12	1
3:	210	1	15:	10	1
4:	192	1	16:	8	1
5:	91	1	17:	7	1
6:	73	1	18:	6	5
7:	65	1	19:	5	4
8:	60	1	20:	4	3
9:	58	1	21:	3	4
10:	42	1	22:	2	3
11:	31	1	23:	1	9
12:	20	1			

Number of "Causal" connective tokens: 1888
Number of "Causal" connective types : 45

Table 8.57 Rank distribution of causal connectives in the English corpus

The top ranks is taken up by the connective "if" with a frequency of 658 that forms 35% of total causals. The next three ranks are occupied by the connectives "because" (247 occurrences representing 13% of causal tokens), "and" (210 occurrences representing 11% of causals) and "so" (192 occurrences, 10% of total causals). The lowest rank holds 9 hapaxes that make up 20% of causal types, though less than 5% of tokens. The type-token ratio in this category is 0.024.

2. Repetitiveness

Indices of repetitiveness of causal connectives in the English corpus are computed as follows:

a. Occurrence rate of causal connectives is 5. This means that in every 5 successive connective tokens one is a causal.

b. Type occurrence rate in this category is 42. This refers to the average number of successive causal tokens there should be before a new type is encountered.

c. General repeat rate is 0.039. This refers to the probability that 2 successive connective tokens - with or without a gap separating them - are any causals.

d. Causal system repeat rate is (66×10^{-4}). That is, the probability that two successive connective tokens turn out to be the same causal type is 66 in 10,000.

e. The gap distribution indicates that the average gap length intervening between two sites occupied by causals is 135 words. The shortest gap length is 0 (two causals running immediately one after another) with a frequency of 14. The longest gap length is 1,714

words having one occurrence. The most frequent gap length is 28 words and has 24 occurrences.

3. Growth

Results of calculations of growth of causal connectives in the English corpus are given in Tables (8.58-59) and plotted in Appendices (122-123).

a. Global Growth: Growth of causal tokens and types are computed within connective intervals, each 5,000 words long. The first interval contains 33 causal tokens and 13 types. The expected numbers are 37 and 15 respectively. The second interval includes 45 causal tokens and 6 new types, raising the number of tokens to 78 and types to 19. Growth fluctuates across the intervals: the highest level of growth is 53 tokens achieved at interval 48, and the lowest level is 24 tokens achieved at interval 43. Growth of

---

Tuldava's (token) richness index:	4.064588
Growth rate of "causal" tokens:	-4.075643
Tuldava's (type) richness index:	0.622379
Growth rate of "causal" types:	14.005765

Actual Text Tokens	Actual Caus.Tokens	Expected Caus.Tokens	Actual Caus.Types	Expected Caus.Types
5000	33	37	13	15
10000	78	75	19	18
15000	113	112	21	20
20000	149	148	22	22
25000	193	185	25	23
30000	238	222	26	25
35000	272	258	26	26
40000	297	295	26	27
45000	348	332	29	28
50000	395	368	31	28
55000	424	404	32	29
60000	454	441	32	30

65000	489	477	32	31
70000	528	514	32	31
75000	556	550	33	32
80000	589	586	34	33
85000	617	623	34	33
90000	642	659	35	34
95000	671	695	36	34
100000	715	731	37	35
105000	743	768	37	35
110000	787	804	37	36
115000	821	840	37	36
120000	854	876	38	36
125000	898	912	38	37
130000	925	948	38	37
135000	951	985	38	38
140000	983	1021	38	38
145000	1014	1057	38	38
150000	1050	1093	38	39
155000	1098	1129	38	39
160000	1150	1165	38	40
165000	1190	1201	39	40
170000	1226	1237	40	40
175000	1277	1273	40	41
180000	1320	1309	41	41
185000	1359	1345	41	41
190000	1398	1381	41	42
195000	1432	1417	42	42
200000	1466	1453	42	42
205000	1492	1489	42	42
210000	1524	1525	42	43
215000	1548	1561	42	43
220000	1586	1597	43	43
225000	1637	1633	43	44
230000	1683	1669	44	44
235000	1723	1705	45	44
240000	1776	1741	45	44
245000	1799	1777	45	45
250000	1834	1813	45	45
255000	1884	1849	45	45
256560	1888	1860	45	45

Extrapolated Text Tokens	Expected Caus. Tokens	Expected Caus. Types
500000	3604	54
750000	5388	61
1000000	7167	66

Table 8.58 Global growth of causal connectives in the English corpus

Tuldava's richness index:	1.044955
"Causal" Connective growth rate:	-1.054693

Actual Connective Tokens	Actual "Causal" Connective Tokens	Expected "Causal" Connective Tokens
500	98	98
1000	202	194
1500	290	290
2000	402	385
2500	489	480
3000	571	574
3500	641	669
4000	728	763
4500	826	857
5000	924	951
5500	995	1045
6000	1098	1139
6500	1210	1233
7000	1333	1327
7500	1434	1420
8000	1536	1514
8500	1642	1607
9000	1768	1701
9500	1873	1794
9596	1888	1812

Extrapolated Connective Tokens	Expected "Causal" Connective Tokens
10000	1888
15000	2819
20000	3747
30000	5597
50000	9281
100000	18441

Table 8.59 Local growth of causal connectives in the English corpus

types show fast growth and early saturation. The 50th type percentile occurs at interval 4 (20,000 words) and the 75th type percentile occurs at interval 15 (75,000 words). After this interval growth slows down considerably. For instance only one new type appears between intervals 24-33 (a total of 50,000 words). The 90th percentile occurs at interval 34 (170,000 words).

Extrapolation of a corpus length of 500,000 words is expected to include 3,604 causal tokens and 54 types. Extrapolation of a million word corpus is expected to include 7,167 causal tokens and 66 types.

b. Local Growth: Growth of causal tokens is also monitored across text intervals each containing 500 connective tokens. The first interval contains 98 tokens, which is the same as the expected number. The next interval comprises 104 more tokens, raising the number to 202. Growth varies from 126 tokens, its highest level, achieved at interval 17 to 71, growth lowest level, at interval 11. Note that at the end of the of corpus, the expected actual number of causals is smaller than the observed one. According to the growth rate, a corpus that is expected to contain 1,888 tokens (the same as the observed number in the present corpus) is estimated to include 10,000 connective tokens. This discrepancy is related to chance fluctuations.

8.9.1.2 Categories of Causal Connectives in English

Distribution of the categories of causal connectives are summarised in Table (8.60) and displayed in Figure (8.15). More details are given in the frequency lists in Appendices (124A-E). The total number of types at this detailed level of categorisation

is 52, marking an increase of 7 types over general causal types. This increase is due to a) the multi-functionality of the connective "lest" (cause, purpose) and "then" (result general, result of a condition); b) multi-use of some connectives whereby they appear as simple and as cores for compound/ correlate connectives: "so" (simple, and core in "so that", "so ... that"), "as" (simple, and core in "as ... so"), "reasons" (core in "for the same reasons" and "for these reasons").

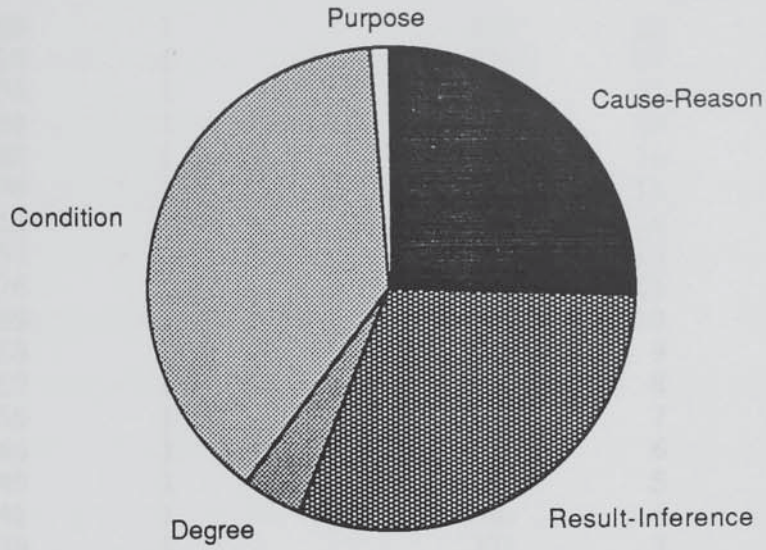
Category	Token	% Token	Type	% Type
Cause	484	25.64	16	30.77
Result-Inference	581	30.77	18	34.62
Condition	724	38.35	11	21.15
Magnitude-Degree	75	3.97	4	7.69
Purpose	24	1.27	3	5.77
Total	1888	100.00	52	100.00

Table 8.60 Distribution of Categories of Causal Connectives in the English Corpus

The largest causal category in terms of tokens is the conditional. It constitutes a set of 724 tokens that correspond to 38% of causal tokens and 7.5% of total connective tokens. The set comprises 11 types, of which 3 are hapaxes. The most frequent connective is "if", which has a frequency of 660, which represents 91% of conditionals, 32% of causals and 7% of total connectives.

The next category in size is result-inference. It has an absolute frequency of 581 representing 31% of causals. The third category is that of cause. It has an absolute frequency of 484 that is equivalent to 26% of causals. The last two categories are relatively smaller: magnitude-degree represents 4% of causals while purpose is equal to only 1%.

Fig. 8.15 Distribution of the Categories of Causal Connectives in the English Corpus



8.9.2 Causal Connectives in Arabic

8.9.2.1 Global Profile

1. Frequency

This is the second largest category of connectives in terms of size (number of tokens) observed in the corpus. It comprises 76 types that have a frequency of 3,778 (see details in the frequency lists in Appendices 125-126). The distribution is displayed in the figures in Appendices (127-128), first within the entire text corpus and then in relation to connective tokens. The rank distribution (Table 8.61) consists of 39 ranks. The first six ranks (each holding one type) achieve 70% of causal token coverage. The top rank is taken by the connective "fa", which has a frequency of 1,160, representing about 31% of causal tokens. The other 5 ranks

Rank	Tokens	Types	Rank	Tokens	Types
1:	1160	1	21:	29	1
2:	418	1	22:	27	1
3:	338	1	23:	21	1
4:	288	1	24:	19	1
5:	268	1	25:	16	1
6:	170	1	26:	15	2
7:	113	1	27:	14	1
8:	97	1	28:	13	1
9:	76	1	29:	11	1
10:	69	1	30:	10	3
11:	65	1	31:	9	1
12:	57	1	32:	8	1
13:	50	1	33:	7	3
14:	46	1	34:	6	1
15:	45	1	35:	5	4
16:	41	1	36:	4	4
17:	39	1	37:	3	4
18:	37	1	38:	2	10
19:	34	1	39:	1	15
20:	30	1			

Number of "Causal" connective tokens: 3778
Number of "Causal" connective types : 76

Table 8.61 Rank distribution of causal connectives in the Arabic corpus

hold the following connectives: "wa" (418 occurrences, equal to 11% of total causal tokens), "'iḏā" (339 occurrences, 9% of causals), "li'anna" (288 occurrences, 7.6% of causals), "li" (268 occurrences, 7% of causals) and "ḏālika ('anna)" (170 occurrences, 4.5% of causal tokens). The lowest rank holds up 16 hapaxes that, while constituting 21% of total types, correspond to only 0.4% of total causal tokens. The type token ratio is 0.02.

2. Repetitiveness

Indices of repetitiveness of causal connectives in Arabic have been computed as follows.

a. Occurrence rate of causal connectives is 4.5. This index means that one causal connective is encountered within an average number of 4.5 connective tokens. This high rate of occurrence is due to the large size of the category.

b. Type occurrence rate within the category is 50. This refers to the average number of causal connective tokens we must have in a text segment before we encounter a new type.

c. General repeat rate is 0.049. This refers to the probability that two successive connectives in a text segment - with or without a gap separating them - turn out to be any causals.

d. Causal system repeat rate is 0.0065. This means that the probability that two successive connectives turn out to be the same causal type is 65 in 10,000.

e. The distribution of gaps separating occurrences of causal connectives in the Arabic corpus suggests that the average gap length is 67 words. The shortest gap length is 0 and has a frequency of 55. The longest gap length runs into 1,059 words. The most frequent gap length is 7 words and has 77 occurrences.

3. Growth

Results of the calculation of growth of causal connectives in Arabic are given in Tables (8.62-63) and displayed in Appendices (129-130).

a. Global Growth: This measure computes growth within text intervals, each 5,000 words long. The first interval contains 63 causal tokens and 15 types. The expected numbers are 66 and 17 respectively. The second interval contains 70 more tokens, raising

---

Tuldava's (token) richness index: 5.003509
Growth rate of "causal" tokens: -4.070068
Tuldava's (type) richness index: 0.833984
Growth rate of "causal" types: -7.094663

Actual Text Tokens	Actual Caus.Tokens	Expected Caus.Tokens	Actual Caus.Types	Expected Caus.Types
5000	63	66	15	17
10000	133	135	20	22
15000	201	206	26	26
20000	279	277	31	30
25000	330	349	35	33
30000	411	421	35	35
35000	487	493	36	38
40000	562	566	40	40
45000	649	639	41	42
50000	732	712	47	44
55000	812	786	51	46
60000	899	859	53	47
65000	992	933	55	49
70000	1120	1007	57	51
75000	1174	1081	58	52
80000	1242	1155	59	54
85000	1306	1229	62	55
90000	1377	1303	62	56
95000	1417	1378	63	58
100000	1479	1452	63	59
105000	1534	1527	63	60
110000	1591	1602	64	61
115000	1665	1676	68	62
120000	1711	1751	68	64
125000	1778	1826	69	65
130000	1840	1901	69	66
135000	1934	1976	69	67
140000	2006	2051	70	68
145000	2091	2126	73	69
150000	2151	2201	74	70
155000	2221	2277	74	71
160000	2302	2352	74	72
165000	2371	2427	74	73
170000	2479	2503	74	74
175000	2560	2578	75	75
180000	2655	2654	75	76
185000	2724	2729	75	77
190000	2793	2805	75	77
195000	2852	2881	75	78
200000	2951	2957	76	79
205000	3023	3032	76	80
210000	3101	3108	76	81

215000	3171	3184	77	82
220000	3231	3260	77	82
225000	3287	3336	78	83
230000	3356	3412	78	84
235000	3439	3488	78	85
240000	3531	3564	78	86
245000	3589	3640	79	86
250000	3657	3716	79	87
255000	3751	3792	79	88
256450	3778	3814	79	88

Extrapolated Text Tokens	Expected Caus. Tokens	Expected Caus. Types
500000	7554	117
750000	11434	140
1000000	15341	158

Table 8.62 Global growth of causal connectives in the Arabic corpus

the token number to 133, and 5 types, putting up the number of types to 20. Growth fluctuates between 128 tokens, growth highest level, at interval 14, to 40, growth lowest level, at interval 19. This considerable variation is due to randomness and chance fluctuation. Extrapolation of a text corpus that comprises 500,000 tokens is expected to contain 7,554 causal tokens and 117 types. Extrapolation of a text corpus of a million words is expected to comprise 15,341 causal tokens and 158 types.

b. Local Growth: This measure monitors the growth of causal tokens within text intervals each containing 500 connectives. The first interval contains 92 causal tokens (the expected number is 94). Growth varies across the intervals from 138 tokens, growth highest level achieved at interval 9, to 82, growth lowest level at interval 26. An extrapolated corpus containing 20,000 connective

Tuldava's richness index:	2.063695
"Causal" Connective growth rate:	-2.010477

Actual Connective Tokens	Actual "Causal" Connective Tokens	Expected "Causal" Connective Tokens
500	92	94
1000	187	198
1500	304	304
2000	407	412
2500	520	521
3000	633	631
3500	742	742
4000	870	853
4500	1008	965
5000	1150	1078
5500	1240	1191
6000	1375	1304
6500	1471	1418
7000	1571	1532
7500	1685	1646
8000	1783	1761
8500	1892	1875
9000	2007	1991
9500	2122	2106
10000	2211	2222
10500	2314	2337
11000	2430	2453
11500	2547	2570
12000	2658	2686
12500	2774	2802
13000	2856	2919
13500	2977	3036
14000	3095	3153
14500	3218	3270
15000	3302	3387
15500	3426	3505
16000	3558	3622
16500	3654	3740
16995	3778	3857

Extrapolated Connective Tokens	Expected "Causal" Connective Tokens
20000	4567
30000	6953
50000	11797
100000	24131

Table 8.63 Calculation of "local" growth of causal connectives in the Arabic corpus

tokens is expected to contain 4,567 causal tokens. An extrapolation of a corpus that contains 100,000 connective tokens is expected to include 24,131.

8.9.2.2 Categories of Causal Connectives in Arabic

Distribution of the categories of causal connectives in the Arabic corpus is summarised in Table (8.64) and displayed in Figure (8.16). Further details are given in Appendices (131A-E). The total number of types at this level of categorisation is 94, showing an increase of 18 types over global causal types. This increase is due

Category	Tokens	% Tokens	Types	% Types
Cause	1259	33.32	33	35.11
Result-Inference	1460	38.65	30	31.91
Condition	554	14.66	12	12.77
Magnitude-Degree	55	1.46	6	6.38
Purpose	450	11.91	13	13.83
Total	3778	100.00	94	100.00

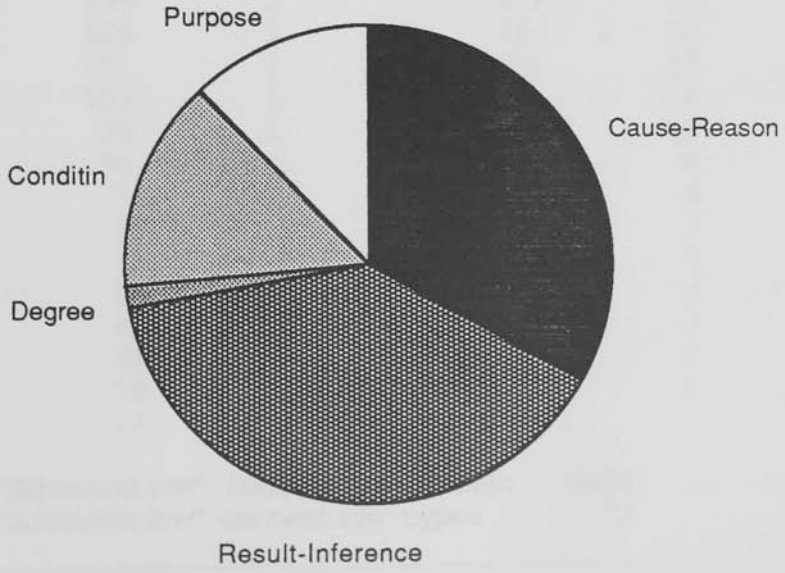
Table 8.64 Distribution of categories of causal connectives in the Arabic corpus

to these two factors: a) the multi-functionality of the following connectives: "fa" (cause, result and consequence of condition), "wa" (cause and result) "li" (result, purpose), "'iḏā" (result, condition), "ḥattā" (magnitude, purpose); b) the multi-use of some connective cores: "sababi" in "bi-sababi" (cause) and "li-hāḏā al-sababi" (result), "wafqan" in "wafqan li" (cause) and "wafqan li" (result), "ḍaw'i" in "<alā-ḍaw'i" (cause) and "fī-ḍaw'i" (result), "'asāsi" in "<alā 'asāsi" (cause) and "<alā hāḏā al-'asāsi" (result) "binā'an" in "binā'an" in "binā'an <alā" (cause) and "binā'an <alay-hi, binā'an <alā ḏālika" (result), "'illā" in "'illā 'iḏā" and "wa

'illā" (both conditionals).

According to the Table, the largest causal category in terms of tokens is the result-inference. It comprises 1,460 tokens, which correspond to 39% of causals, and contains 33 types, of which the most frequent are "wa" and "fa". The next category, in a descending order of frequency, is that of cause. It consists of 1,259 tokens and comprises 33 types, of which the most frequent two are "fa" and "li'anna". The categories of condition and purpose come next in frequency. "Condition" connectives comprises a set of 554 tokens and 12 types while purpose connectives consist of 450 tokens and have 13 types. The smallest category in size is that of magnitude-degree: 55 tokens (1.5% of causals) and 6 types.

Fig. 8.16 Distribution of the Categories of Causal Connectives in the Arabic Corpus



8.10 Adversatives Connectives

8.10.1 Adversative Connectives in English

8.10.1.1 Global Profile

Adversative connectives in the English corpus constitute the largest category in terms of tokens. The category comprises 72 types that have a total frequency of 2,404. Details of the frequency distribution are given in Appendices (132-133) and displayed in the figures in Appendices (134-135). The rank distribution (Table 8.65) consists of 29 ranks. The first 5 ranks, each occupied by one type, achieve approximately 70% of total adversative token coverage. At the top rank comes the connective "but" with a frequency of 1,167, representing 48.5% of total adversative tokens and equal to 12% of total connective tokens in

Rank	Tokens	Types	Rank	Tokens	Types
1:	1167	1	16:	16	1
2:	190	1	17:	15	1
3:	125	1	18:	14	1
4:	101	1	19:	13	1
5:	100	1	20:	11	1
6:	99	1	21:	10	1
7:	86	1	22:	9	1
8:	73	1	23:	8	3
9:	66	1	24:	7	3
10:	33	1	25:	5	7
11:	32	1	26:	4	5
12:	22	1	27:	3	5
13:	21	1	28:	2	7
14:	19	1	29:	1	19
15:	17	2			

Number of "Adversative" connective tokens: 2404
Number of "Adversative" connective types : 72

Table 8.65 Rank distribution of adversative connectives in the English corpus

the corpus. The next 4 ranks hold the following four connectives: "however" (190 occurrences equivalent to 7.87% of total adversatives), "though" (125 occurrences, 5.2% of adversatives), "although" (101 occurrences, 4% of adversatives), "yet" (100 occurrences, 4% of adversatives). The lowest rank is occupied by 19 hapaxes that, while they constitute 26% of types, achieve less than 0.8% of adversative token coverage. The type-token ratio in this category is 0.03.

2. Repetitiveness

Indices of repetitiveness of adversative connectives in the English corpus have been computed as follows:

a. Occurrence rate of adversatives is 4. This high rate of occurrence is due to the large size of category.

b. Type occurrence rate is 33, slightly slower than the general connective occurrence rate. This is again due to both the enormity of the size of category in terms of tokens and the small number of types.

c. General repeat rate is 0.063. This refers to the probability that two successive connective tokens - with or without a gap separating them - are any adversatives.

d. Adversative system repeat rate is 0.016. This refers to the probability that two successive connectives turn out to be the same type.

e. Calculation of gap distribution shows that the average distance length is approximately 106 words. The shortest and most frequent distance is 0, which has 46 occurrences. The longest distance is 1,182 words, which occurs once.

3. Growth

Results of global and local growth of adversative connectives are offered in Tables (8.66-67) respectively and plotted in Appendices (136-137).

a. Global Growth: This measure monitors the growth of adversative tokens and types within successive text intervals of 5,000 words each. The first interval contains 42 tokens and 18 types, close to the expected numbers, which are 40 and 18. The second interval contains 43 more tokens, but only 5 new types. Growth of tokens varies within the categories from 59, growth highest level, achieved at interval 14, to 18 at interval 36. Types grow fast during the first 12 intervals and then drops to a minimum level. This is evident in the places where the 50th, 75th and 90th percentile types occur. The 50th percentile type occurs early in the corpus: within interval 7 (35,000 words); the 75th percentile type occurs within

---

Tuldava's (token) richness index:	5.096293
Growth rate of "adversative" tokens:	-5.042536
Tuldava's (type) richness index:	0.697904
Growth rate of "adversative" types:	-25.002468

Actual Text Tokens	Actual Adv.Tokens	Expected Adv.Tokens	Actual Adv.Types	Expected Adv.Types
5000	42	40	18	18
10000	85	83	23	23
15000	113	127	26	27
20000	164	172	30	30
25000	205	218	31	33
30000	255	264	33	35
35000	310	310	37	37
40000	367	356	40	39
45000	417	403	43	40
50000	465	450	43	42

55000	506	497	46	44
60000	550	544	47	45
65000	592	592	49	46
70000	647	639	50	48
75000	706	687	52	49
80000	753	735	53	50
85000	804	782	53	51
90000	860	830	55	52
95000	907	879	56	53
100000	960	927	57	54
105000	999	975	57	55
110000	1046	1023	60	56
115000	1089	1072	60	57
120000	1145	1120	61	58
125000	1173	1169	62	59
130000	1231	1218	63	60
135000	1270	1266	63	61
140000	1325	1315	63	61
145000	1371	1364	63	62
150000	1418	1413	65	63
155000	1470	1462	66	64
160000	1521	1511	67	65
165000	1576	1560	67	65
170000	1622	1609	67	66
175000	1677	1658	67	67
180000	1704	1707	67	67
185000	1751	1757	68	68
190000	1803	1806	68	69
195000	1845	1855	68	69
200000	1883	1905	69	70
205000	1934	1954	69	71
210000	1981	2004	71	71
215000	2028	2053	71	72
220000	2080	2103	71	73
225000	2119	2152	71	73
230000	2164	2202	71	74
235000	2208	2252	71	74
240000	2251	2302	71	75
245000	2289	2351	71	76
250000	2339	2401	71	76
255000	2391	2451	72	77
256560	2404	2466	73	77

Extrapolated Text Tokens	Expected Adv. Tokens	Expected Adv. Types
500000	4924	98
750000	7489	114
1000000	10082	127

Table 8.66 Global growth of adversative connectives in the English corpus

Tuldava's richness index:	2.072027
"Adversative" Connective growth rate:	-2.008199

Actual Connective Tokens	Actual "Adversative" Connective Tokens	Expected "Adversative" Connective Tokens
500	104	105
1000	212	222
1500	355	342
2000	478	464
2500	587	587
3000	728	712
3500	854	838
4000	975	964
4500	1093	1092
5000	1222	1219
5500	1341	1348
6000	1471	1477
6500	1599	1606
7000	1722	1736
7500	1847	1866
8000	1998	1997
8500	2123	2128
9000	2243	2259
9500	2371	2391
9596	2404	2416

Extrapolated Connective Tokens	Expected "Adversative" Connective Tokens
10000	2522
15000	3853
20000	5200
30000	7931
50000	13479
100000	27634

Table 8.67 Calculation of "local" growth of adversative connectives in the English corpus

interval 18 (90,000) words; the 90th type percentile occurs within interval 30 (150,000 words). After that, types gradually reach saturation.

An extrapolated corpus of 200,000 words is expected to contain 4,924 adversative tokens and 98 types. An extrapolated corpus of a million words is expected to include 10,082 adversative tokens and 127 types.

b. Local Growth: This measure computes the growth of adversative tokens within text intervals each containing 500 connectives. The first interval contains 104 adversative tokens (the expected number is 105). The next interval contains 108 more tokens, raising the number to 212. Growth fluctuates between 151 tokens, growth highest level, at interval 16, and 104, growth lowest level, at the first interval. Extrapolation of a text corpus that contains 10,000 connective tokens is expected to contain 2,522 adversatives. An extrapolated corpus containing 100,000 connective tokens is expected to include 27,634.

8.10.1.2 Categories of Adversative Connectives in English

Distribution of the categories of adversative connectives is summarised in Table (8.68) and plotted in Figure (8.17). Further details concerning the frequency of each category is given in

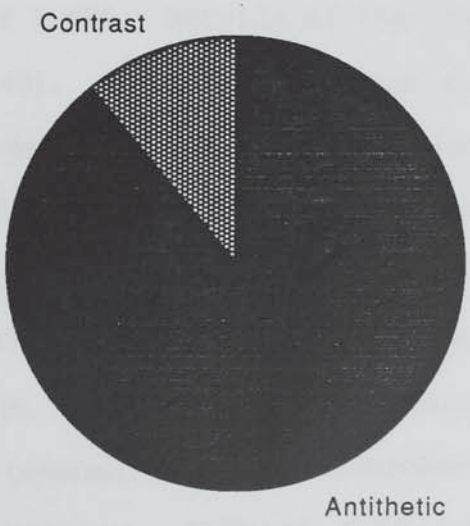
Category	Tokens	% Tokens	Types	% Types
Antithetic	2149	89.39	49	65.33
Contrast	255	10.61	26	34.67
Total	2404	100.00	75	100.00

Table 8.68 Distribution of categories of adversative connectives in the English corpus.

Appendices (138A-B). The total number of types at this level of categorisation is 75, giving an increase of 3 types over global adversative type count. This increase is effected by such factors as the multi-use of "even", "but", serving as simple connectives and cores for compound ones ("even-if", "but then").

Antithetic connectives constitute the bigger of the two categories: 2,149 tokens, corresponding to 89% of adversative tokens and 22% of global connective token count. The set comprises 49 types (equal to 65% of adversative types), of which 12 are hapaxes. The most frequent connective is "but", which has a frequency of 1,156, standing for 54% of antithetic tokens, 48% of total adversative tokens and 12% of global connective tokens. Next in the

Fig. 8.17 Distribution of the Categories of Adversative Connectives in the English Corpus



order of frequency come "however" (189 occurrences, 9% of antithetic tokens), "though" (125 occurrences), "although" (101), "yet" (100), "even" as a core in "even if" and "even though" (74), adversative "and" (73), "of course" (66), and "whatever" (33).

The next category of adversative connectives is that of contrast. The category consists of 255 tokens, representing 10% of adversatives and approximately 3% of total connective count, and comprises 26 types of which 8 are hapaxes. The most frequent connective is "while", having a frequency of 99 that represents 39% of contrast tokens and 4% of total adversative token count. The next 5 most frequent connectives are "instead" (32 occurrences), "on the other (hand)" (22), "rather" (17), "whereas" (15), "on the contrary" (11).

8.10.2 Adversative Connectives in Arabic

8.10.2.1 Global Profile

1. Frequency

This category of connectives consists of 53 types that have a total frequency of 1,692. Details of the frequency are given in Appendices (139-140). The distributions are presented in the figures in Appendices (141-142). The rank distribution (Table 8.69) is composed of 29 ranks. The first six ranks (holding one type each) have a frequency of 1,060 that achieves 63% of adversative token coverage. These six ranks are occupied by the connectives: "lākinna" (top rank, a frequency of 310 corresponding to 18% of total adversative tokens), "wa" (306 occurrences, equivalent to 18% of adversative tokens), "lākin" (169 occurrences, 10% of adversatives), "'innamā" (101 occurrences, 6% of adversatives),

Rank	Tokens	Types	Rank	Tokens	Types
1:	310	1	16:	17	3
2:	306	1	17:	16	2
3:	169	1	18:	14	1
4:	101	1	19:	13	2
5:	98	1	20:	12	1
6:	76	1	21:	11	1
7:	52	1	22:	9	1
8:	47	2	23:	8	1
9:	44	1	24:	7	1
10:	40	2	25:	6	2
11:	34	1	26:	4	1
12:	33	1	27:	3	3
13:	29	1	28:	2	9
14:	25	1	29:	1	8
15:	20	1			

Number of "Adversative" connective tokens: 1692
Number of "Adversative" connective types : 53

Table 8.69 Rank distribution of adversative connectives in the Arabic corpus

"illā 'anna" (98 occurrences, 5.8% of total adversatives), and "bal" (76 occurrences, 4.5% of total adversatives). The lowest rank contains 8 hapaxes. The type-token ratio of this category is 0.031.

2. Repetitiveness

Indices of repetitiveness of adversative connectives in the Arabic corpus are computed as follows:

a. Occurrence rate of adversative connectives is 10. This index refers to the average number of successive connective tokens we may have in a text segment before we encounter an adversative token.

b. Type occurrence rate within the category is 32, referring to the number of adversative tokens that a text segment may have before

a new type appears.

c. General repeat rate is 0.01. This is the probability that two successive connective tokens - with or without a gap - are both adversatives.

d. Adversative system repeat rate is (9×10^{-4}). This is the probability that two successive connectives turn out to be the same adversative type.

e. The gap distribution indicates that the average distance length is 150 words. The shortest gap is 0 (no intervening words between two successive occurrences of adversative tokens) having a frequency of 6. The longest gap length is 2,478. The most frequent gaps are 11 and 13 words, each having a frequency of 17.

3. Growth

Results of calculation of global and local growth of adversative connectives in Arabic are given in Tables (8.70-71) respectively and displayed in Appendices (143-144).

a. Global Growth: Growth of adversative tokens and types are first monitored in successive text intervals of 5,000 words each. The first interval contains 49 tokens and 14 types. The expected numbers are 45 and 15 respectively. The next interval includes 42 more tokens, which raises the number of tokens to 91, and 5 new types, which increases types to 19. Growth varies considerably from 58, the highest level, achieved at interval 50, to only 8, the lowest level at interval 18.

Growth of types is fast during the first few intervals. The 50th percentile type occurs during the interval 5 (25,000 words). After that growth is slow. The 75th type percentile appears at interval

---

Tuldava's (token) richness index:	2.097253
Growth rate of "adversative tokens:	-3.078455
Tuldava's (type) richness index:	0.650552
Growth rate of "adversative" types:	38.000876

Actual Text Tokens	Actual Adv.Tokens	Expected Adv.Tokens	Actual Adv.Types	Expected Adv.Types
5000	49	45	14	15
10000	91	83	19	19
15000	116	119	23	22
20000	154	154	25	24
25000	193	188	27	26
30000	217	222	27	27
35000	245	255	30	29
40000	273	288	30	30
45000	320	320	32	31
50000	364	352	36	32
55000	398	384	36	33
60000	424	415	36	34
65000	462	446	38	35
70000	490	477	40	36
75000	517	508	40	37
80000	550	538	41	37
85000	565	569	42	38
90000	573	599	42	39
95000	594	629	42	39
100000	618	659	42	40
105000	639	689	42	41
110000	659	719	42	41
115000	684	748	42	42
120000	715	778	42	42
125000	746	807	43	43
130000	771	836	44	44
135000	797	866	44	44
140000	826	895	44	45
145000	854	924	45	45
150000	890	953	45	46
155000	928	981	45	46
160000	980	1010	46	46
165000	1020	1039	46	47
170000	1059	1067	48	47
175000	1105	1096	48	48
180000	1147	1124	48	48
185000	1174	1153	48	49
190000	1216	1181	48	49
195000	1248	1209	48	49
200000	1292	1238	49	50
205000	1327	1266	49	50
210000	1368	1294	50	51

215000	1392	1322	51	51
220000	1417	1350	51	51
225000	1441	1378	51	52
230000	1477	1406	51	52
235000	1523	1434	51	52
240000	1551	1461	51	53
245000	1587	1489	52	53
250000	1645	1517	53	53
255000	1674	1545	53	54
256450	1692	1553	53	54

Extrapolated Text Tokens	Expected Adv. Tokens	Expected Adv. Types
500000	2858	66
750000	4145	75
1000000	5398	82

Table 8.70 Global growth of adversative connectives in the Arabic Corpus

14 (70,000 words). Signs of saturation start to appear after this interval, and, indeed, within the next 100,000 words, the number of new types do not exceed more than 15% of adversative types in the corpus. The 90th type percentile occurs within interval 34 (170,000 words).

An extrapolated corpus of 500,000 is expected to contain 2,858 adversative tokens and 66 types. A corpus of a million words is expected to contain 5,398 adversative tokens and 82 types.

b. Local Growth: Growth of adversatives is computed in successive text intervals, each containing 500 connectives. The first interval contains 65 adversative tokens (the expected number is 59). The highest level of growth is 76 tokens achieved at intervals 9 and 33. The lowest level of growth is 24 achieved at interval 12. Note that

Tuldava's richness index:	1.034151
"Adversative" Connective growth rate:	-1.079082

Actual Connective Tokens	Actual "Adversative" Connective Tokens	Expected "Adversative" Connective Tokens
500	65	59
1000	108	113
1500	171	164
2000	217	214
2500	256	264
3000	296	312
3500	372	360
4000	420	408
4500	466	456
5000	503	503
5500	549	550
6000	573	596
6500	616	642
7000	648	688
7500	696	734
8000	747	780
8500	787	825
9000	826	871
9500	864	916
10000	926	961
10500	983	1006
11000	1042	1051
11500	1095	1096
12000	1150	1140
12500	1193	1185
13000	1248	1229
13500	1307	1273
14000	1368	1318
14500	1414	1362
15000	1455	1406
15500	1519	1450
16000	1565	1494
16500	1641	1537
16995	1692	1581

Extrapolated Connective Tokens	Expected "Adversative" Connective Tokens
20000	1842
30000	2697
50000	4365
100000	8403

Table 8.71 Calculation of "local" growth of adversative connectives in the Arabic corpus

at the end of the corpus the observed number of tokens is larger than the expected one, 1,692 tokens compared to 1,581. This is an indication of the extent of fluctuation in growth across the corpus. Extrapolation of a corpus containing 20,000 connective tokens is expected to include 1,842 adversative tokens. A corpus that contains 100,000 tokens is expected to have 8,403.

8.10.2.2 Categories of Adversative Connectives in Arabic

Distribution of the categories of adversative connectives is summarised in Table (8.72) and plotted in Figure (8.18). More details are given in Appendices (145A-B). The total number of types in the two adversative categories is 58, yielding an increase of 5 types over global adversative types. This results from the use of

Category	Token	% Token	Type	% Type
Antithetic	1359	80.32	37	63.79
Contrast	333	19.68	21	36.21
Total	1692	100.00	58	100.00

Table 8.72 Distribution of categories of adversative connectives in the Arabic corpus

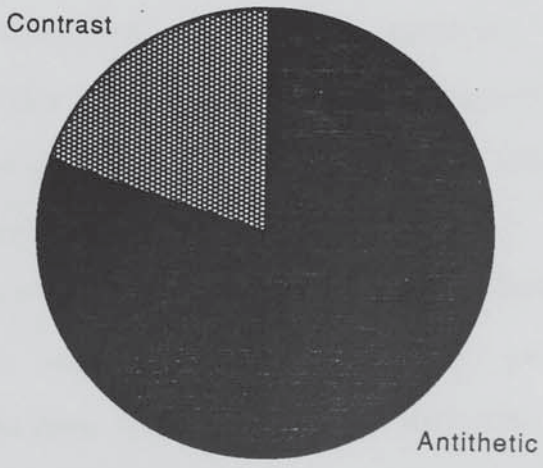
"'idā" to denote two functions and the use of "'illā" and "'in" as simple connectives and as cores in the multi-word connectives "'illā 'anna" and "'in ... 'am".

Antithetic connectives constitute the larger of the two categories. They consist of 1,359 tokens, equivalent to 80% of

adversative tokens and 8% of total connective token coverage. This category comprises 37 types, of which 7 are hapaxes. The most frequent connective is "lākinna" [but] (310 occurrences, equivalent to 23% of antithetic connectives, 18% of adversatives).

Connectives of contrast comprise 333 tokens that constitute 20% of adversatives and 2% of total connective token coverage. This category includes 21 types, of which only 2 are hapaxes. Apart from adversative "wa", which occurs 108 times, the next in frequency are "baynamā" [while] (34 occurrences) and exceptive/adversative "'illā" [except that] (30 occurrences).

Fig. 8.18 Distribution of the Categories of Adversative Connectives in the Arabic Corpus



8.11 Conclusion

This chapter has been devoted to the second part of the calculus of observations on connectives. Specifically, we have been concerned with outlining a statistical account of each functional category of connectives in both corpora.

The starting point for the account is a short statistical preview of functionality in both corpora, where the categories within each corpus are compared to each other in size and proportion. Each category is, then, selected individually and three main statistical features are discussed: a) frequency of distribution and some related measures; b) measurement of some aspects of repetitiveness: system occurrence rate, type occurrence rate, general and system repeat rates and gap distribution; c) measurements of "general" and "local" growth.

The next step in the analysis is a brief discussion of the main statistical features of the various subcategories that constitute each main category. This account has been kept to a minimum so as not to burden the chapter unnecessarily.

What emerges from the statistical profiles is not only the variation that exists in the deployment of connectives across the two languages, but also the wide variation in the share that each functional category assumes within the individual corpus. In each corpus, categories differ in the type or token coverage they achieve, in their repetitiveness and their growth rate. The differences, we believe, are commensurate with the textual role that members of a category play. In other words, the differences in the distribution of categories represent variations in the way connectivity is patterned: some categories have a more dominant

influence on connectivity than others.

To pursue this argument, one might be tempted in this context to formulate generalisations for the two languages. However, it would be too early to attempt to pass judgements in regard to patterns of connectivity as yet, despite the fact that the evidence we have obtained is neither slender nor deficient. Rather, we would like first to explore the inter-lingual, as well as inter-categorial, variations with more detail and see how we can exploit the evidence that we have gathered in making qualified confirmation of the size and quality of the differential patterns of connectivity. This major task is left to the next chapter.

Footnotes to Chapter 8

(1) Full forms of tagged connectives are given in Appendices 14 (English) and 15 (Arabic). A detailed consideration of the quantitative characteristics of full forms is considered superfluous for the purpose of this project.