COHESION IN THE SPOKEN ENGLISH DISCOURSE OF ARABS

By

YOUSIF ELHINDI

Bachelor of Arts
University of Khartoum
Khartoum, Sudan
1975

Master of Arts
Sheffield University
Sheffield, United Kingdom
1980

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December 1995

COHESION IN THE SPOKEN ENGLISH

DISCOURSE OF ARABS

Thesis Approved:

Carol Lynn Moder

Thesis Advisor

Janet B. Pittman

Susan Garzon

Larry Hochhaus

Thomas C. Collins

Dean of the Graduate College

ACKNOWLEDGEMENTS

Praise be to Allah for granting me the ability and the perseverance to go through the hardship of graduate school. I have been blessed with a number of supporters who helped me through this endeavor. It is hard to express my gratitude to all of them in this limited space. First, I would like to express my sincere thanks and deep appreciation to my adviser Dr. Carol Moder for her superior guidance and continued inspiration. I have benefited a lot from her exemplary teaching and excellent supervision. My deep appreciation also goes to Dr. Gene Halleck who generously spent a lot of her valuable time and effort to do the interviews and who introduced me to the interesting field of OPI. I would also like to thank the other two members on my advisory committee Drs. Susan Garzon and Larry Hochhaus for their encouragement and suggestions.

My thanks are also extended to the brothers and sisters who agreed to be interviewed for this study and to all members of the Islamic Society of Stillwater who have made the past five years spiritually rewarding. I am also grateful to my friends Nadine and Paul Woods for rating the interviews and helping me with the computer and Michael Dordick for his encouragement.

I would also like to express my gratefulness to my friends Becky and Steve Damron who made our stay in Stillwater a pleasant one. My thanks are also extended to the faculty and staff of the English Department, especially Drs. L. Leff, P. Rollins, and R. Batteiger for their help. I am also thankful to Shirley Bechtel, Dale Mclaughlin, and Sherl Holesko for helping me with the tables.

I am also indebted to my uncle Yousif Hamad for his help and encouragement. I would also like to extend my sincere thanks and gratefulness to my brother Siddig for his

continuous financial and moral support. This work could not have been completed without his generosity and help. My appreciation also goes to my brothers Abdul and Ali and my nephew Elamin for their support and encouragement.

Finally, I express my love and gratitude to my wife Huda for her unwavering support and encouragement throughout the years and to our loved sons and daughters: Ahmed, Mohamed, Raghda, and Elaf.

TABLE OF CONTENTS

Chapter Page

APPENDICES, cont.

LIST OF TABLES

LIST OF TRANSCRIPTION SYMBOLS

Capital letters	:	beginning of turns, proper nouns, and after question marks
. .	:	short pause
. . .	:	long pause
[ext. pause]	:	extended pause
" "	:	foreign words
!	:	exclamation/animated tone
-	:	interruption
( )	:	mispronunciation e.g. (skeen) for "scene"
.	:	end of turn (marked by a short pause)
bold	:	abnormal emphasis

Chapter I

Introduction

The modern trend in linguistic research is to investigate language that is used in authentic situations. The basis for this shift in emphasis is that the primary function of language is communication. Since communication is collaborative and interactive, the present focus of research is on how people manage to achieve their communicative goals through language. One of the important features that characterize meaningful discourse is cohesion which refers to the syntactic, lexical, and pragmatic ties that hold discourse together and make it cohesive.

These cohesive ties, however, do not guarantee the meaningfulness of language since it is possible to construct an instance of discourse that is cohesive but whose parts are not logically related to each other. In other words, cohesion, which refers to the way the different parts of an instance of language stick together, does not imply coherence, or the logical connectedness of that instance of discourse. This argument had often been used by the critics of cohesion research who pointed out that a piece of discourse could have a substantial number of cohesive ties but these ties do not necessarily guarantee that this discourse is logically interrelated, or coherent. Proponents of the investigation of cohesion, however, are not unaware of this fact. Halliday and Hasan (1976), who proposed the most comprehensive taxonomy for the investigation of cohesion in English,

repeatedly reiterate throughout their book that the context of situation, including para-linguistic and extra-linguistic features should always be taken into consideration.

Despite this criticism, a tremendous body of research using a variety of methods has been carried out in the field of cohesion. Most of this research has focused on written language as we shall see in the following chapter. This was the consequence of the importance of the written medium of language in academic circles and the convenience of analyzing written rather than spoken discourse. During the past few years, however, interest in spoken discourse has generated research that looks into spoken discourse. Although studies investigating both written and spoken language have often drawn conflicting conclusions as to the significance of formal cohesive ties in discourse, the consensus is that cohesive relations in discourse are important phenomena that merit investigation.

Most of the research investigating cohesion in written discourse has used Halliday and Hasan's taxonomy. The reason for this is that Halliday and Hasan's model is designed to examine cohesion between sentences. And since the 'sentence' is a grammatical unit that lends itself more to written than spoken language, researchers have found the taxonomy more readily applicable to written rather than spoken discourse. Halliday and Hasan themselves exclusively illustrated their model by applying it to written texts, especially Alice in Wonderland.

This trend, however, started to change with the collapse of the Chomskyan paradigm and the emergence of disciplines that investigate how people use spoken language in authentic situations, e.g. sociolinguistics, discourse analysis, and pragmatics. As a result

of this other models of cohesion that were specifically designed to investigate spoken discourse started to surface. One of these is that proposed by Schiffrin (1987b) which defines various linguistic forms and explains how they are employed by discourse participants to achieve a number of communicative and pragmatic goals.

One of the genres investigated by this research has been academic nonnative speakers' (NNS) discourse. This was sparked by the growing number of nonnative teaching assistants at American universities. As students, parents, and administrators started to express their concern about how these teaching assistants - despite their good knowledge of the subject matter they teach - lacked the oral proficiency necessary for classroom teaching. Consequently, research investigating the characteristics of academic nonnative discourse started to develop.

Although this type of research is important, it investigates a nonnative discourse genre that is completely different from other kinds of nonnative discourse. First, lectures are different from other discourse genres with respect to their format which makes them less collaborative than other interactive genres. Since most of the talking is done by the lecturer, she is the one who delivers almost the whole discourse unless, of course, a question is asked during the lecture. Second, academic discourse is planned, hence, teachers usually walk into the classroom with a prepared agenda specifying what they are going to talk about and in what order. Third, because of the high load of new information in lectures, this genre is characterized by the use of certain linguistic forms that are meant to help students organize and, hence, comprehend the lecture.

Since this genre is quite unique, it is important to investigate other genres of NNSs' discourse for a number of reasons. First, the findings of academic discourse research can not be generalized to other types of discourse. Second, nonacademic NNS discourse is an important genre that is widely used by NNSs in their encounter with both native speakers (NS) and NNSs from different language backgrounds who use English as a lingua franca. Third, although a few studies have investigated cohesion in nonnative discourse, none, to my knowledge, has used a comprehensive cohesion model to investigate the discourse of NNSs from different proficiency levels.

This dissertation is an attempt to carry out such an investigation. It uses a cohesion model that combines Halliday and Hasan's (1976) and Schifrin's (1987b) models to look into cohesion in the spoken English discourse of Arabs from different proficiency levels.

The dissertation falls into five chapters. This chapter - Chapter One - discusses the significance of the phenomenon investigated, i.e. cohesion. It introduces the main models proposed to study cohesion and argues that research in this field has, to a large extent, focused on written rather than spoken language. The chapter also argues that nonacademic nonnative spoken discourse is an important genre that has not received its due share of research.

Chapter Two reviews some of the research that looked into cohesion. The variety of the studies in this chapter are meant to reflect the different approaches used to study cohesion and give the reader an overview of the diversity of the genres that have been investigated. The diversity of the approaches together with the variety of the models

applied to investigate cohesion - it is assumed - is partly responsible for the conflicting findings reached by some of these studies.

Chapter Three uses an elaborate model of cohesion to thoroughly investigate this phenomenon. The Chapter also describes the method for this study and argues for investigating cohesion in the spoken discourse of NNS from different proficiency levels. This includes an account of how the subjects were selected, an explanation of the procedure employed, and an illustration of how the different types of cohesion were coded.

Chapter Four reports the results of the analysis for all six cohesive relations. It shows the tables for the distribution of these ties and explains the trends reflected by the distribution and reports the statistical measures used to determine the significance of the difference between the three proficiency levels in using these cohesive ties. The Chapter also explains how the quality and not the quantity of cohesion ties marks levels of proficiency.

Finally, Chapter Five presents the conclusions of this dissertation. The Chapter also discusses the implications of this study to future research and to a number of disciplines to which oral proficiency is relevant.

Chapter II

Review of Literature

Cohesion, which may be defined as the grammatical and/or lexical relationships that exist between various parts of spoken or written texts, is an essential feature of discourse. Despite the different approaches that have been proposed for the study of cohesion (Halliday and Hasan, 1976; Beaman, 1984; Derrick-Mescua and Gumca, 1985; and Planalp et al. (1987) to name just a few,) this vital phenomenon remains an elusive one. This is partly due to the lack of consensus among researchers as to what constitutes cohesion, and to the argument that formal cohesive markers, though important, do not necessarily guarantee cohesion.

In this chapter, some of the major models that have been proposed for the study of cohesion will be reviewed. In addition, the application of these models will be evaluated by reviewing some of the studies that have employed them to analyze cohesion.

The most extensive taxonomy proposed for the study of cohesion is Halliday and Hasan's (1976). They define cohesion as the semantic relations within a text that give a text its meaning. Halliday and Hasan believe cohesion is what distinguishes meaningful discourse from disconnected, meaningless sentences. Cohesion, they point out, takes place when the decoding of a particular component in the discourse can not be carried out

without turning to another element for its interpretation. When this condition is satisfied, "a relation of cohesion is set up, and the two elements, the presupposing and the presupposed, are thereby at least potentially integrated into a text" (p. 4). Since cohesion is a semantic relation between the components of the discourse, Halliday and Hasan argue, the "presupposing" and the "presupposed" need not be structurally related to each other. Moreover, they argue since the elements of the sentence are cohesive by virtue of the structural relationships that hold them together, it is more important to investigate intrasentential cohesion. Hence, they limit their proposed taxonomy of cohesion to the study of cohesive ties between sentences. Using examples from written English texts, they state that cohesion is achieved through ties that fall into five categories: reference, substitution, ellipsis, conjunction, and lexical cohesion.

The first category in Halliday and Hasan's taxonomy is *reference*. They assert all languages have items whose interpretation is dependent on making reference elsewhere. When the reference is made within the textual context, it is called endophora, and when it is made outside the text - in the situational context - it is called exophora. An example of exophoric reference is:

[1:1] That was a flagrant foul by Charles Oakley.

where *that* refers to something outside the linguistic context, and an example of endophoric reference is:

[1:2] Olajuwon averaged over thirty points in Houston's sweep of Orlando. He also won his second consecutive NBA finals MVP.

where *he* in the second sentence makes reference to *Olajuwon* in the first.

Further, Halliday and Hasan classify endophora into anaphora and cataphora. Anaphoric reference requires the retrieval of a presupposed item in the preceding text. An example of this is [1:2] where the presupposing *he* points back to the presupposed *Olajuwon* in the preceding text. On the other hand, cataphoric reference constitutes the retrieval of a presupposed element from the following text:

[1:3] They dismissed their fifth head coach in two years .. the LA Clippers.

where *they* points forward to the presupposed *LA Clippers*.

In addition to their classification by directionality, Halliday and Hasan identify three types of reference in English: personal, demonstrative, and comparative. Personal reference is further divided into three subclasses which are personal pronouns, possessive determiners, or possessive adjectives, and possessive pronouns. They state there is no cover term in English grammar that includes these various subclasses and they propose *person* arguing although the category includes non-personal reference items, it is easier to remember and it is less fuzzy than most grammatical terms.

Furthermore, Halliday and Hasan point out personal pronouns have an important semantic role in the communication process because they identify discourse participants. Among the personal pronouns, they argue, "only the third person is inherently cohesive, in that a third person form typically refers anaphorically to a preceding item in the text" (p. 48). The first person "I/we" and the second person "you", on the other hand, usually make reference to the speaker/hearer and writer/reader in spoken and written language respectively.

The second type of cohesion in Halliday and Hasan's model of analysis is *substitution* which denotes instances where an item is replaced by another e.g.,

[1:4] Judge Ito dismissed a fourth juror today. Rumor has it that another one is being investigated.

Here *one* in the second sentence substitutes for *juror* in the first. Halliday and Hasan debate that since "the substitute item has the same structural function as that for which it substitutes" the cohesive relation of *substitution* "is a relation on the lexicogrammatical level, the level of grammar and vocabulary, or linguistic form" (p. 89).

They identify three types of substitution: nominal, verbal, and clausal. Nominal substitution involves the substitution of one/*ones* for a form functioning as head of a nominal group e.g.,

[1:5] I like this new headset better. It is more comfortable than the old one.

Here *one* in the second sentence substitutes for *headset* in the first. The second type of substitution is nominal substitution which is fulfilled by *same* e.g.,

[1:6] President Clinton threatened to increase the tariff on Japanese luxury cars. The same was considered last year.

same in the second sentence is referential in the sense that it points back to increasing the tariff on Japanese luxury cars; whereas in:

[1:7] A: I will have a whopper and medium fries.

B: I will have the same, please.

same functions as a nominal substitute.

The second type of substitution which is verbal substitution is realized by *do* in English. Halliday and Hasan assert that verbal and nominal substitutions have a number of similarities since both the nominal and the verbal groups in English have some similarities in their structure. Each has a head and an optional modifier or modifiers. And in either one it is possible to substitute for the head, with *one* in the nominal group and *do* in the verbal group. However, the major difference between the two is that *one* substitutes only for a noun (see 1:4), whereas *do* substitutes for either a verb as in:

[1:8] John McEnroe never won the French Open. He might have done it had he practiced more on clay surfaces.

or a verb in addition to other clause elements as in:

[1:9] A: Is it true that tuition fees at OSU are going to increase by 7% next fall?

B: I am afraid so.

In this example *so* substitutes for the clause "that tuition fees at OSU are going to increase by 7%".

The third type of cohesion is *ellipsis*, which Halliday and Hasan argue, is very similar to substitution. They define ellipsis as "substitution by zero" (p. 143) arguing that *one* and *do* are used in substitution to mark the place of a presupposed item. In ellipsis, on the other hand, nothing is inserted to fill the place of the presupposed information; however, it is understood from the previous text. They divide this type of cohesion into three categories: nominal, verbal, and clausal ellipsis.

Nominal ellipsis takes place when a modifier in the nominal group is upgraded to the status of *head* of the nominal group:

[1:10] Two NBA coaches were fired during the regular season, and two during the playoffs.

Two in the first clause is a numerative modifying the head of the nominal group *coaches*. In the second clause; however, *two* has been promoted to the status of *head* to stand for or presuppose the elliptical *NBA coaches*.

The second type of ellipsis is verbal ellipsis which occurs within the verbal group. The most common type of verbal ellipsis is what Halliday and Hasan call lexical ellipsis where an elliptical modal operator is used and a lexical verb is left out:

[1:11] A: Is Bob Dole going to run for president in 1996?

B: He might.

This is an instance of lexical verbal ellipsis where an elliptical verbal operator *might* presupposes a lexical verb *run*. The second most common type of verbal ellipsis, operator ellipsis, occurs when an elliptical lexical verb presupposes an operator. This mainly happens in closely related discourse sequences like questions and answers:

[1:12] A: Have you been watching the Cowboys on TV?

B: No, transcribing interviews.

In this example, two operators *have* and *been* were omitted whereas the lexical verb remained intact.

The fourth type of cohesion is *conjunction*. Unlike other cohesive relations, Halliday and Hasan assert, conjunctive forms are indirectly cohesive because of their inherent semantic properties. They indicate how the meaning of one element in the text should be interpreted in relationship to another e.g.,

[1: 13] Ibrahim lost all the data for his research. So he won't graduate this semester.

In this example, the conjunction *so* denotes a causal relationship between the two propositions in the text: *losing the data* and *not graduating*.

Halliday and Hasan identify four basic types of relations expressed by conjunctions: additive, adversative, causal, and temporal. These can be realized by a variety of conjunctions that belong to different grammatical categories including coordinating conjunctions e.g., *but, so, then*; adverbs e.g., *nevertheless, moreover, however*; and prepositional phrases e.g., *on the other hand* and *at the same time*. Moreover, they distinguish between external and internal conjunctive functions which can be illustrated by:

[1:14] He was collecting data for his dissertation. At the same time, he was writing the review of literature.

where the conjunctive element *at the same time* is a temporal one showing that the two events expressed in the text took place at the same time; hence, it is an instance of an external use of temporal conjunction. An example of the internal type is:

[1:15] Pete Samparas is an excellent grass court player. At the same time, he does not play well on clay courts.

In this example, the same temporal conjunction *at the same time* is an internal one showing that the speaker is trying to express two ideas at the same time; hence, the temporal relation expressed by the conjunction is internal to the speech situation.

The fifth and final category in Halliday and Hasan's model of cohesion analysis is *lexical cohesion* which is the repetition of semantically related vocabulary items. This repetition creates the presupposition that the full relevance of the repeated word or phrase

is only interpretable with reference to the first use. Unlike other types of cohesion which are realized by specific linguistic forms, any "lexical item may enter into a cohesive relation, but by itself it carries no indication whether it is functioning cohesively or not" (p. 288). This is determined by taking into consideration the "particular collocational environment that has been built up in the course of the creation of the text and that will provide the context within which the item will be incarnated on this particular occasion. This environment determines the 'instantial meaning', or text meaning, of the item, a meaning which is unique to each specific instance" (p. 289).

Further, they classify lexical cohesion into reiteration and collocation. Reiteration is realized by: a) repetition, or using the same item; b) synonyms or near synonyms; c) superordinates; and d) using a general word. Collocation, the second type of lexical cohesion, "is achieved through the association of lexical items that regularly co-occur" (p. 284). Let us illustrate this type of cohesion by investigating some of the lexical ties in the following excerpt:

> [1:16] (1) The **game** is not out of his system. (2) We can understand that. (3) No **athlete's** love for his **sport** has ever been greater than **Magic Johnson's** affection for **basketball**, and a bond that strong cannot be broken simply by holding a retirement press conference or watching a number raised to the rafters. (4) **Magic** cannot walk away that easily, nor should anyone expect him to. (5) But there comes a time when a clean break is necessary, when holding on becomes undignified, even a bit unseemly. (6) It is getting to that point with **Johnson**, who, despite his success as a businessman and as vice-president and part owner of the

> **Los Angeles Lakers**, cannot seem to give up his vision of himself as a **player**. (7) **He** has been hinting for weeks that he wants to return to the courts, and last week **Laker's** executive vice president Jerry West finally took the hint, saying **Los Angeles** is interested in a Magic reappearing act. (*Sports Illustrated*, July 17, 1995).

Some instances of lexical reiteration in this passage include repeating the same word e.g., *Magic*, employing synonyms e.g., *love/affection*, and using a superordinate e.g., *sport* that includes the hyponyms *game* and *basketball*. Furthermore, the excerpt consists of collocations such as antonyms e.g., *break/hold on* and several lexical chains e.g., *game .. sport .. basketball .. player,* and *retirement .. press conference .. number .. rafters.*

Concluding the review of Halliday and Hasan's model for cohesion analysis, I believe it is important to reiterate a point that they frequently emphasize throughout their book, especially the first chapter. They stress that the mere presence of the cohesive ties that they classify and explain does not guarantee cohesion. What should be taken into consideration, they point out, is not only the formal linguistic ties, but also the context of situation including all para-linguistic and extra-linguistic factors. They rightly argue:

> one can construct passages which seem to hang together in the situational-semantic sense, but fail as texts because they lack cohesion, so also one can construct passages which are beautifully cohesive but which fail as texts because they lack consistency of register - there is no continuity of meaning in relation to the situation (p. 23).

Most of those who have criticized Halliday and Hasan (as we shall see below) have focused on how they ignored the situational context. Ironically, the examples that these

critics cited to support their point of view are of the kind mentioned in the last part of the quotation above.

Halliday and Hasan's approach has also been criticized for applying a model of analysis that is limited to written language. This, some discourse analysts argue (Carrel, 1982), does not take into account important phenomena such as contribution of participants to the understanding of discourse (Charolles, 1983), distributional limitations of cohesive devices in different discourse genres (Ehrlich, 1988), and the relations between text and the real world (Foster, 1984). Despite these limitations, Halliday and Hasan's approach has the advantage of proposing a comprehensive, systematic model of cohesion analysis that paved the way for other approaches to follow.

Halliday and Hasan's model has been applied in a large number of studies. This is partly due to the fact that it is the most comprehensive taxonomy because it proposes the tools necessary to account for grammatical, semantic, and lexical cohesion. However, most of the research that has applied this system of analysis (as we shall see below) has been restricted to written texts rather than spoken discourse. The reason behind this is twofold. First, the model lends itself to the analysis of written texts because cohesive ties are investigated between sentences. And the sentence has traditionally been regarded as a written unit that is clearly defined by the rules of sentential grammar and orthographic conventions. This was demonstrated by the examples given by Halliday and Hasan which have all been taken from written texts. The second reason that this model appealed to many investigators is that it is a multifaceted taxonomy that enabled researchers to investigate the type of cohesion they thought suitable for their research purposes. Since

most of the research in this area has been carried out during the 70's and early 80's, it was consistent with the trend at that time which emphasized the use of written texts as objects of linguistic analysis.

As the interest in spoken discourse increased with the collapse of the Chomskyan paradigm and the growing body of literature in the fields of sociolinguistics, discourse analysis, and pragmatics in the 80's, this genre started to receive the due attention of researchers and, consequently, models of coherence in spoken discourse started to emerge.

One of these was proposed by Schiffrin (1985, 1987a, 1987b). This model is proposed for the analysis and interpretation of spoken English and it focuses on what she calls local coherence, or coherence between adjacent units in discourse. This, she believes, is accomplished by employing *discourse markers* that speakers use to achieve, not only semantic, but also pragmatic goals. Schiffrin does not give a concrete definition of discourse markers arguing that she is being deliberately vague because these forms occur "at the boundaries of units as different as tone groups, sentences, actions, verses, and so on" (Schiffrin, 1987, p. 36). The operational definition that she gives is that discourse markers are "**sequentially dependent** elements which bracket units of talk" (p. 31). Despite this limitation which arises from the difficulty of defining speech units, Schiffrin identifies a number of discourse markers applying them to the analysis of authentic conversation explaining how they are used by participants to achieve coherence through accomplishing a variety of communicative goals.

One of these discourse markers is *well* which, Schiffrin (1985) argues, is different from all other discourse markers because it does not have inherent semantic or structural attributes. Its meaning derives from the context in which it occurs. Schiffrin analyzed the distribution of *well* in conversation obtained through sociolinguistic interviews. She distinguished between occurrences of *well* in adjacency pairs i.e. questions/answers, request/compliance sequences; and occurrences of *well* which cannot be explained in terms of adjacency pairs.

Examining the distribution of *well* in answers to wh-questions and yes/no questions, Schiffrin found that *well* precedes answers to yes/no questions 10% of the time, on the other hand *well* preceded answers to wh-questions 21% of the time. She attributed this to the fact that yes/no questions delimit the upcoming answer to either affirmative or negative response, hence *well* is not regarded as a coherent marker for answers in these contexts. Her examination of other types of questions that are similar to yes/no questions in delimiting responses e.g., tag questions, confirmed the findings above. In answering wh-questions respondents used *well* 14% of the time when they supplied the wh-information - sought by the question - in their answers. When wh-information was not provided, respondents used *well* 56% of the time. This, Schiffrin believes, proves that *well* precedes answers that diverge from the options given by the question.

Schiffrin (1985) also examined the role of *well* in response to requests. She suggests that in response to requests for action it is more likely for *well* to be used in responses that indicate non-compliance to requests. As for requests for information, she gives - among others - the following example:

[1:17] Ira: And I have been working for the federal government ever since. Thir-
thirty-six years hhhh.

Debby: So you must liked them as an employer then.

Ira: Well, I like my job now. (Schiffrin, 1985, p. 652).

Since *now* restricts Ira's statement to the present time, Schiffrin argues that *well* marks a non-compliance response to the request for confirmation in the preceding statement.

As for occurrences of *well* outside adjacency pairs, Schiffrin identifies four different contexts. The first is when *well* is used to mark requests. This category includes requests for clarification when a speaker needs more information about a previous utterance before s/he could make a response. The second category includes using *well* to mark response "whose referent is larger than, and contains discourse prior to, the immediately prior utterance" (p. 657). The third context includes occurrences of *well* as a marker of self repairs and speaker's monitoring of her talk. Finally, the fourth category includes the occurrence of *well* as a marker of reported speech. Since reported speech is different from the discourse at hand with regard to *setting* and *participants*, *well* is used to indicate an orientation shift.

Schiffrin concludes her paper by stating that *well* is a marker that makes it possible for the participants to achieve various communicative goals at the different levels of discourse in their attempt to achieve coherence. She emphasizes the need for future research that examines the distribution of different markers in different discourse genres. Such research, she asserts, would enable us to understand how coherence is achieved in conversational discourse and other discourse genres.

In a subsequent study that attempted to account for the distribution and significance of a number of discourse markers, Schiffrin (1987b) emphasizes the importance of discourse contexts and their influence on the production and interpretation of utterances. She defines local context as the adjacent utterances within a conversation. Then she distinguishes between two kinds of analyses: sequential accountability which refers to the explanation of coherence within a text, and distributional accountability which explains the distribution of certain elements in certain discourse genres but not in others. Schiffrin associates sequential accountability with qualitative analysis which examines the characteristics of a limited body of data, and associates distributional accountability with quantitative analysis which examines generalities within a large body of data. Moreover, she stresses the need for the combination of both approaches to achieve empirical discourse analysis.

Schiffrin (1987b) proposes a model of discourse analysis that consists of five domains. These domains are: information state, participation framework, ideational structure, action structure, and exchange structure. The last two domains are non-linguistic pragmatic structures which include turns and speech acts respectively. *Ideational structures* are semantic structures that include: 1) *cohesive* relations, when an interpretation of a clause element requires information from prior discourse, 2) *topic relations*, what the discourse is about, and 3) *functional relations,* which refers to the roles of ideas within a text. *Participation framework* refers to the speaker-hearer relationship and what is being said and done. Finally *information state* includes

knowledge and metaknowledge, that is, speaker/hearer's knowledge and their knowledge about each other's knowledge.

Schiffrin states that discourse markers, which are utterance initial elements, enable speakers to understand each other's intentions. Hence, they are essential guidelines to discourse analysts because they serve as indices to local contexts. She identifies 11 discourse markers: *oh, well, and, so, now, then, I mean, but, because, y'know* and *or*. Schiffrin adds that a single discourse marker can simultaneously point to more than one local context. *Well*, for instance, can be an answer to a question (an index to the exchange structure), at the same time it can indicate the insufficiency of that answer (an index to the ideational structure). However, each discourse marker has a primary function in a particular local context.

Schiffrin concludes her study by reiterating the contribution of discourse markers to achieving coherence. These devices, she stresses, can accomplish a variety of communicative goals which include denoting shifts in orientation, anchoring speakers' participation to previous and upcoming discourse, managing discourse time, and marking the knowledge and metaknowledge of the participants in a speech situation. These functions, Schiffrin adds, integrate the various facets of the discourse and, thus, make it cohesive.

A more recent study by Schiffrin (1990) determined the distinction between *anaphorics*, or forms whose referent lies in the linguistic text, and *deictics*, which index certain aspects in the contextual situation, is not always clear-cut. She analyzed the occurrences of the temporal adverb *then* in a variety of contexts to illustrate her point.

First, Schiffrin discussed the function of *then* as a deictic temporal form stating that it indexes the time of a specific event which does not coincide with the speaking time. This function contrasts with *now* which has a proximal meaning. Both forms, however, indicate relative temporal distance and do not specify an absolute point in time.

Furthermore, Schiffrin discussed the anaphoric function of *then* distinguishing between two functions:

Q: (1) Will you see Joan when you go to New York?

A: (2) Yes, I will see her then.

(3) Then I will see her again when I go to Chicago (Schiffrin, 1990, p. 254).

Then, which occupies a clause-final position in A (2) denotes two overlapping events: *going to New York* and *seeing Joan*. The second type is successive *then* which is exemplified by A (3). In this example, *then* occurs in a clause-final position and it denotes two successive events: "seeing Joan in New York" and "seeing her in Chicago".

These two functions of *then* as a deictic (A2) and an anaphoric (A3), Schiffrin (1990) states, are not always clear-cut. She argued there are instances of anaphoric *then* where the event denoted by anaphora is in the situational context rather than the linguistic text. She illustrated this through a remark made by her son when he was two and a half years old. While he was looking at a photograph of himself taken on his first birthday he remarked:

I was laughing then (Schiffrin, 1990, p. 257).

Schiffrin argues *then*, which seems to be an anaphoric in this example, denotes an event *laughing* that is contexually rather than texually situated.

On the other hand, Schiffrin states, it is also possible for deictic *then* to function textually. She gives the following example from a television show where a chef demonstrating a recipe said:

A: 1. I peel the apple,

2. Then I put it in the processor ... (p. 260).

Schiffrin argues "since each event is concurrent with the moment of speech, ... *then* does not have a distal interpretation at all" (p. 260). To further illustrate her point, she pointed out proximal *now* could be substituted for distal *then* without changing the interpretation:

A: (3) I peel the apple,

(4) Now I put it in the processor ... (p.260).

Consequently, Schiffrin determined *then* is simultaneously functioning as a deictic and an anaphoric. This led her to state anaphora depends on deixis since temporal relationships in linguistic texts are only understood through the interpretation of the relationships between text and context. She concluded "language is not 'situated in' in social interaction ... rather, language is a 'part of' social interaction such that language and social interaction both create, and are created from, each other" (p. 265).

Schiffrin's model is a pioneering approach that looks into a variety of words and phrases, or discourse markers, from a functional perspective. She investigates how these markers are used by speech participants to achieve a variety of communicative goals. Schiffrin's approach also has the advantage of providing the tools to analyze coherence in spoken discourse. This has not been taken into consideration by Halliday and Hasan whose model was specifically tailored for written texts. Moreover, Schiffrin's approach is

not restricted by the limitations of sentential grammar terminology. She has succeeded in identifying and accounting for a grammatically diverse group of words, *conjunctions*, *adverbs*, *interjections*, and *phrases*.

However, Schiffrin's compilation of discourse markers is by no means comprehensive. There are still many discourse markers that need to be identified and analyzed, e.g. *anyway* and *okay*. Moreover, Schiffrin's model which divides the relations between discourse units into five structures (see above) has a number of fuzzy areas. First, Schiffrin claims that each discourse marker can simultaneously function at more than one level; however, each marker has one primary function. The problem with this claim is that it is not clear how each marker is assigned a primary function because Schiffrin does not clearly define the criteria for making such a decision. Second, her definition of discourse markers is vague. This seems to be the result of trying to identify their location in relation to a speech unit, a task that has puzzled researchers for too long. Third, the theoretical perspective on which she tries to anchor her analytical model is too ambitious because she attempts to account for all features of discourse including the informational content, discourse exchange structure, speech acts, and knowledge and metaknowledge of discourse participants. This seems to be necessitated by the conventions and expectations of academic research which values analytical models that are based on theoretical assumptions.

Despite the shortcomings, Schiffrin's approach is a valuable contribution to the field. It provides the tools for looking into how participants in discourse use a variety of markers in a cooperative manner that facilitates discourse comprehension.

Another model of cohesion analysis that attempted to account for every aspect of discourse was proposed by Charolles (1983). He criticizes Halliday and Hasan's distinction between coherence as related to the situational context, and cohesion as a set of formal ties that give discourse its texture and are intended to aid the receiver in her interpretation of the discourse. Charolles argues that the distinction is not clear because formal cohesive ties are only valuable if they are perceived in a specific situation. As a solution, he proposes an approach that attempts to explain cohesion by accounting for the *message* and the *setting*. In addition to these it also identifies all the traits of the participants. His procedural approach also attempts to account for the processing of the data by considering, not only the qualities that are necessary for formal discourse coherence, but also the goals of the communicative act and how discourse is perceived and processed by the participants at the different stages of communication. Charolles' model draws on findings in sociolinguistics, e.g. Hyme's (1974) communicative competence, that advocate the accommodation of a variety of situational factors in the explanation of discourse. Although Charolles' approach is an ambitious one that looks very comprehensive and appealing in theory, he admits "there are still a lot of things left undone in the system" (p. 85) especially in the area of how interlocutors process and interpret discourse. This also explains why no attempt was made by the researcher to apply his model to an actual analysis of discourse.

Another model of discourse that has attempted to account for cognitive processing of linguistic forms has been developed by Givon (1992). Givon argues that textual grammar, despite its advantages that include explaining the relationship between grammar and the

situational context, has failed to accommodate into its parameters the cognitive contribution of discourse participants in coding and interpreting texts. Givon emphasizes that grammatical referential devices do not by themselves achieve coherence; however, they serve as indexes that identify and activate mental storage areas into which incoming information is deposited.

Givon's view is synonymous with the current interactive trends in linguistics that emphasize the role of participants in the creation and interpretation of discourse. This trend states that *meaning* is not inherent in texts, rather it is achieved through the interaction of the writer/reader with the written text, or the speaker/hearer with the spoken discourse. Thus, such approaches emphasize the cognitive aspect of the communicative process by investigating the cognitive processes responsible for coding and decoding discourse.

Givon states that since discourse is interactive, participants have to cognitively accommodate each other's perspective. This, he adds, is achieved through specific devices that make the information accessible and comprehensible to the hearer. These communication facilitating devices include the organization of discourse into *given* information, or information supposed by the speaker as known to the hearer, and *new* information, or information the speaker assumes as unknown to the hearer. Givon hypothesizes that the cognitive function of old information is to "furnish the *address* or *label* for the *storage locus* ('file') in the episodic memory" (p. 9). Grammatically, new information is realized by nouns which function as discourse topics and are realized by "the grammatical subjects and objects of clauses" (p. 9).

Givon then proceeds to look into two measurable aspects of discourse which are *referential accessibility* and *thematic importance*. And he attempts to account for them from a cognitive perspective. *Referential accessibility*, Givon states, is dependent upon the deictics of the situational context, the linguistic context, and cultural knowledge. On the other hand, *thematic importance* guides the hearer to activate storage areas that consist of important topics and deactivate unimportant topic files.

Reviewing research that has investigated these two phenomena, Givon argues that although this methodology has improved our understanding of texts, it has failed to account for the cognitive processing in discourse. To overcome this shortcoming, Givon proposes a model that explains how referential devices are utilized by decoders to process discourse. He postulates that the cognitive counterparts of *thematic importance* and *referential accessibility* are: a) attentional activation; b) episodic and permanent memory, respectively. Givon then hypothesizes four processes triggered in the *attentional system* by grammatical forms of reference. These are:

a. keeping the current file open

b. closing the current file

c. opening a new file that is

(i) opened for the first time

(ii) reopened after being inactive for some time (p. 23).

Givon hypothesizes that each of these processes is associated with certain grammatical and cognitive considerations. The first process, keeping a current file open, is grammatically signaled by the use of unstressed pronouns and zero anaphora. Givon calls

this the "grammatical code quantity principle" arguing that "information that is already activated requires the smallest amount of code" (p. 25). These minimal code units, Givon adds, are much more frequent in discourse than maximal code units, or stressed pronouns and noun phrases, and they require less attention and cognitive effort.

The second process, process (b) which is closing the current file, always precedes (c i) and (c ii) since only one file can remain active in the episodic memory. The usage of an indefinite noun phrase (NP) emits the cognitive instruction that a decision concerning the activation or deactivation of a file should be put off until its importance is determined. If the NP is unimportant, then a new file should not be opened nor should an existing inactive file be activated. Thus, the new information should be deposited in the file that is currently open. On the other hand, if the NP is important, a new file into which subsequent information is deposited should be opened and activated (c ii). Givon, however, does not account for differences between entities that have been recently active and entities that have been activated earlier in the discourse.

Givon argues that definiteness is not a grammatical property per se since some languages, like Mandarin Chinese, mark indefinites and leave definites unmarked. Givon adds that definite NPs require more cognitive effort since the source of definiteness should first be determined before they are processed. This is achieved through scanning for the antecedent in three sources: a) the situational context, or deictics; b) the cultural context, or permanent memory; and c) the textual context, or episodic memory. After the definiteness source is determined by locating the antecedent in one of these three sources,

the existing inactive file should be reopened and activated. And all subsequent information should be deposited into that file.

Since this process of activating an existing file is the most cognitively demanding of all others, Givon discusses it in more detail. First, he specifies the cognitive correlates of the three domains of definiteness arguing that situational context corresponds to the "mental image of speech situation" (p. 31), the cultural context to the permanent memory, and the textual context to the episodic memory. When a search for the antecedent of a definite NP is triggered by a particular grammatical form, Givon postulates, the speech situation mental model will be scanned first. If the antecedent is not available in that domain, a search will be carried out in the permanent memory. Finally, the episodic memory will be scanned if the first two searches were unsuccessful.

Givon concludes his paper arguing that although the hypotheses formulated in this study were based on research on narrative discourse (Schank and Abelson, 1977; Mandler, 1978; Van Dijk and Kintsch, 1983), the findings could be applied to collaborative genres. He states that there is no evidence at the present as to the nature of how discourse is stored in the minds of the participants, nor what the structure of that discourse information is. However, he proposes a number of assumptions relevant to these issues. He postulates that episodic memory is where collaborative discourse is stored. Moreover, Givon hypothesizes that when the discourse is uncooperative, and thus less coherent, it is stored as two independent entities in each participant's episodic memory. On the other hand, collaborative discourse, which is more coherent, is stored as a single unit by each participant with each one of them clearly marking and indexing her turns or contributions.

These hypotheses seem to be based on the assumption that in less coherent collaborative discourse the contributions of the participants do not mesh together. As a result of this mismatch, speakers store each contribution as a separate entity. On the other hand, cohesive collaborative discourse is stored as one unit by each speaker because each participant's contribution match the other's.

Givon's research differs from other cohesion/coherence studies in a number of ways. First, like all cognitive grammar studies, it attempts to define grammatical forms from a cognitive perspective by accounting for how different linguistic forms contribute to the interpretation of discourse through triggering a variety of cognitive processes. This is of particular importance in the field of cohesion/coherence since most research in this area has focused on cohesion, or the formal linguistic ties that link units of discourse, ignoring coherence, or the logical sequencing of discourse units. Such research is also vital because the semantic aspect of language, or comprehension processes, have always been either unresearched, or described in some vague way in terms of respondent reaction. However, since Givon's findings are based on research in the field of narrative discourse (see above), they need to be supported by research that looks into the distribution of these grammatical forms in collaborative discourse since collaborative and narrative discourse are two different genres that might constrain the distribution of grammatical forms differently.

The last model of cohesion analysis to be reviewed in this section is that proposed by Hoey (1991). Hoey criticizes Halliday and Hasan claiming that although they have acknowledged lexical coherence and incorporated it in their taxonomy, they have failed to

develop the proper model. He adds that Halliday and Hasan's categories for lexical cohesion, which are reiteration and collocation, are vague and insufficient to account for this vital aspect of cohesion. Hoey argues that reiteration in Halliday and Hasan's model includes a variety of categories that consist of items with different semantic relations while collocation has not been used in the proper sense of the word which is accounting for the relationship between items "that appear with greater than random probability" (p. 7).

As an alternative, Hoey (1991) proposes a system of lexical analysis that explains lexical repetitions in texts. He classifies repetitions into four types: a) simple lexical repetition, or the repetition of a lexical item without altering its form; b) complex lexical repetition, or repetition of lexical items that share a lexical morpheme; c) paraphrases; and d) a subcategory that includes personal pronouns *he*, *she*, *it*, and *they*; demonstrative pronouns; substitutes *one*, *do*, and *so*; and a small group of items that include *other*, *same*, *different*, and *similar*.

These relationships, Hoey argues, are all instances of repetition that create bonds between sentences in texts. These bonds, or links between lexical items in different sentences, are significant because they index sentences that are informationally loaded. Hoey argues that a sentence which has more lexical bonds with other sentences is more crucial to the text in which it occurs than sentences with few bonds.

Hoey proposes a system of lexical analysis that includes a matrix of relationships. The lexical bonds for each sentence are represented by a two-digit coordinate, a right number for the bonds with previous text and a left number for bonds with subsequent text. Hoey states that a high left coordinate indicates the sentence is "topic initiating" while a high

right coordinate shows that the sentence is "topic concluding". Hoey argues that sentences with low coordinates, or marginal sentences, can be deleted without affecting the cohesion of the text. On the other hand sentences with high coordinates serve as good summaries for the texts in which they occur. Hoey also employs visual representations, or nets, using lines to connect lexical bonds between sentences. Below is an illustration of Hoey's model using the *Magic Johnson* text (example 1:16). The lines connect instances of bonds in the text. Only the important instances of bonds that have the highest weight or importance have been marked (see Hoey, p. 83).

[1:18] The **game** is not out of his system. We can understand that. No **athlete's** love for his **sport** has ever been greater than **Magic Johnson's** affection for **basketball**, and a bond that strong cannot be broken simply by holding a retirement press conference or watching a number raised to the rafters. **Magic** cannot walk away that easily, nor should anyone expect him to. But there comes a time when a clean break is necessary, when holding on becomes undignified, even a bit unseemly. It is getting to that point with **Johnson**, who, despite his success as a businessman and as vice-president and part owner of the **Los Angeles Lakers**, cannot seem to give up his vision of himself as a **player**. **He** has been hinting for weeks that he wants to return to the courts, and last week **Laker's** executive vice president Jerry West finally took the hint, saying **Los Angeles** is interested in a Magic reappearing act. (*Sports Illustrated*, July 17, 1995).

Hoey's lexical net, exemplified by the above passage, indicates that sentences (6) and (7) are the most important sentences in the text with 7 and 8 ties or connections, respectively. Sentences (2) and (5), on the other hand, are the most marginal or least important because neither of them has any ties. As for the coordinates, being the final sentence, sentence (7) has a 0 right coordinate; however, it has an 8 left coordinate, which is the highest of all sentences. This supports Hoey's claim that concluding sentences have the most lexical ties with previous text. However, the topic initiating sentence, sentence (1), has a 2 right coordinate while sentence (6) has a 6 right coordinate, the highest of all sentences. This does not support Hoey's claim that topic initiating sentences have the greatest number of ties with subsequent text.

There are many shortcomings in Hoey's proposed model. First, the criteria that he specifies for determining what constitutes a bond is fuzzy. In the paragraph about "drugging grizzly bears" which he uses to illustrate his analytical model, *animals* and *bears* are regarded as an instance of a lexical bond, whereas *biologists* and *scientists* are not treated as one. Second, in order for him to prove the claim that marginal sentences can be dropped without affecting the cohesion of the text, he manipulated the text by changing the structure of one of the sentences. Third, Hoey admits that there are pairs of sentences that cohere via means other than those accounted for by his model. An example of such a type of cohesion is illustrated by sentence (2) above which has two instances of cohesive ties that are not accounted for by Hoey's model. These are the personal pronoun *we* and the deictic *that*. He attributes this to the inadequate "techniques of lexical analysis" (p. 160).

Although the five models of cohesion discussed above set out to explain the same phenomenon, they have more differences than similarities. The times at which these models were published reveal the linguistic trend that has colored the theoretical assumptions on which these models were based. Halliday and Hasan (1976) was published at a time when the majority of linguistic analyses used written texts as objects of linguistic investigation at the expense of spoken discourse. Consequently, most of those who applied this model looked into the distribution of cohesive ties in written language. Although Halliday and Hasan have frequently reiterated that their model only investigates one aspect of coherence which is texual cohesion - the other being the situational context - they have been heavily criticized for ignoring the functional aspect of language. Despite the shortcomings of Halliday and Hasan's taxonomy (see above), it has proved to be a comprehensive model of analysis that has been used in numerous studies as diverse as ESL, native oral discourse and composition, genre constraints on cohesive ties selectivity, reading comprehension, nonnative academic discourse, and literary analysis. This is mainly due to the explicitness and comprehensiveness of the taxonomy which enabled researchers to apply the aspects that best suit their research purposes.

The second model of cohesion analysis (Charolles, 1983) is an ambitious theoretical account that defines what aspects of discourse should be analyzed when coherence is investigated. Charolles, however, does not attempt to do an actual analysis of discourse admitting that his system is not developed enough to carry out such a task.

The third model of cohesion analysis (Schiffrin, 1987b) was specifically developed to account for a small group of linguistic forms. These forms, or discourse markers, enable

speakers to accomplish a variety of communicative tasks as reviewed above. Schiffrin's approach is completely different from Halliday and Hasan's in a number of ways. First, as previously mentioned, the model is devised to investigate spoken rather than written language. This was consistent with the trend during the late 1980s which witnessed a growing interest in sociolinguisics, discourse analysis, and pragmatics. All these disciplines emphasize the importance of spoken language in authentic situations. Second, the discourse markers are identified by a functional criterion that looks into their role in discourse, rather than a grammatical criterion which assigns them to a specific category. Third, Schiffin has attempted to investigate a particular aspect of cohesion through an in-depth analysis of a number of linguistic forms that have not been previously considered as functionally similar. Halliday and Hasan, on the other hand, have proposed a comprehensive model of analysis to account for a wide range of cohesive ties.

The fourth model of analysis, proposed by Hoey (1991), attempts to investigate patterns of lexical cohesion in texts. Although Hoey's discussion of lexical repetition in the first three chapters gives the impression that only lexical items are accounted for in his analysis, his model includes other types of cohesive ties like *substitution* and *ellipsis*. However, Hoey argues that these are the least important in defining cohesion whereas lexical repetitions are the most important. Hoey's model is similar to Halliday and Hasan's in that it includes all their categories of cohesive ties except *conjunction*. Nevertheless, there are more differences than similarities between the two approaches. These are: 1) Halliday and Hasan's is a much more clearly defined approach that can easily be applied to identify cohesive ties, 2) Hoey argues that cohesive ties vary in weight, or

importance, and he identifies a hierarchy of links with lexical repetitions and paraphrases as the most important and substitution and references as the least important, and 3) Hoey's approach that assigns number coordinates for repetition in sentences seems to lend itself to corpus and computational linguistics techniques. Although these techniques make the analysis of a large body of data possible, they muddy the distinction between linguistic forms that are often given the same numerical value despite their differences. The system, for example, considers lexical relationships like *synonymy*, *antonymy*, and *hyponymy* as instances of reiteration and does not account for their semantic differences.

The fifth and final model (Givon, 1992) is quite different from all other approaches in a number of ways. First, it is the only model that attempts to provide a cognitive explanation of the processing of cohesive ties, i.e. grammatical references, in discourse. Second, the theoretical assumption of Givon's approach, which is consistent with the current trend in linguistics, is that texts do not constitute meaningfulness by themselves. The interaction of the discourse participants - speakers/hearers with spoken discourse and writers/readers with written discourse - is what creates the meaning of discourse. Third, in so doing, Givon's approach attempts to bridge the gap between cohesion and coherence by investigating the role of grammatical forms in comprehension.

Of the five models of cohesion reviewed above, Halliday and Hasan's has been the approach most frequently applied to the investigation of cohesion. One of the studies that applied this model was carried out by Witte and Faigley (1981) who, in an attempt to test the validity of Halliday and Hasan's theory, applied it to the analyses of essays written by 10 freshmen at the University of Texas. Out of 90 essays they selected 5 that were given

the highest score by four raters. Their analyses confirmed that high-rated essays contained a cohesive tie every 3.2 words, compared to a cohesive tie every 4.9 words in the low-rated essays. They also concluded that the mean number of cohesive ties per T-unit in the high-rated essays was 5.2, compared to 2.4 per T-unit in the low rated essays. This led them to conclude highly rated essays have more frequent cohesive ties than low-rated essays. However, they argued that despite the usefulness of cohesive analysis as a research tool, it is not by itself sufficient to gauge good writing quality. Consequently, they suggest certain outside-text factors have to be considered when writing is evaluated. These include writer's purpose, discourse medium, and audience's knowledge of the subject.

Another study that investigated cohesion in written texts employing some of Halliday and Hasan's techniques was carried out by Johnson (1987). Using Halliday and Hasan's categories of *reference*, *conjunction*, and *lexical* cohesion, Johnson's analysis determined that "good" compositions written in the Malay language consisted of more semantic ties, *reiteration* and *collocation*, than "weak" compositions. On the other hand, "good" compositions written in English by native speakers had more syntactic ties, *reference* and *conjunction*, than "weak" compositions. Johnson made the tentative conclusion that these differences are either the result of the constraints of topic - Malay students wrote descriptive essays while native English speakers wrote expository ones - or they are an indication of the cross-cultural differences as to what constitutes a well written essay.

Beside its application to the investigation of cohesion in written English by native and nonnative speakers, Halliday and Hasan's model has been used to determine the

constraints of genre on the selectivity of cohesive ties. Smith and Frawley (1983) investigated *conjunctive cohesion* in four different genres of English texts from the Brown English Corpus: fiction, journalism, religion, and science. They argue that the study of cross-genre conjunctive cohesion is important for two reasons. The first is that conjunctions mark relationships between constituents that span beyond the clause boundaries unlike, for instance, prepositions which mark relations within the clause. Second, they argue most research on conjunctive cohesion has focused on narration, ignoring other genres of texts. This, they believe, is inadequate since it ignores the constraints genre types might have on the frequency, distribution, and types of conjunctions. Such a cross-genre approach, the writers assert, enables us to understand the logic of the discourse.

Smith and Frawley investigated the distribution of conjunctions in 16,000-word texts of religion, science, journalism, and fiction. Their findings indicated that both the type and frequency of conjunctions were different across the four genres. Coordinating conjunctions amounted to 66.4% of the total number of conjunctions in fiction, 59.4% of the conjunctions in science, 55.4% in religion, and 54.7% in journalism. Subordinating conjunctions, on the other hand, were 33.6%, 40.6%, 44.6%, and 45.3% respectively. They state that these findings which indicate the number of coordinating conjunctions exceed that of subordinating conjunctions in all four genres refute the common belief that sophisticated types of writing employ more subordination than coordination. The researchers also concluded there was no correlation between the frequency of conjunctions and mean sentence length. They argue that although it is quite reasonable to

expect genres using long sentences to use more conjunctions, this turned out to be true only for religion. Both science and journalism texts whose mean sentence length was 21.4 and 23.8 words respectively - compared to the mean sentence length of the whole corpus which was 19.3 - used conjunctions less frequently.

Besides investigating the frequency of conjunctions, Smith and Frawley also looked into the semantic categories of the conjunctions across the four genres. Such an analysis, they argue, would help explain the rhetorical and semantic structure of these different text types. Using Halliday and Hasan's four categories of conjunctions: additive, adversative, causal, and temporal, they concluded that 80% of the conjunctions in the fiction genre were adversative and additive. They attributed this " to the seriation of events intuitively characteristic of the genre" (p. 366). As for journalism, they found out it was marked by the employment of temporal conjunctions. They attributed this characteristic to the importance to the timing of events in this type of writing. The researchers also determined that religion was characterized by the frequent use of the additive *nor* and the adversative *yet*. This similarity between religion and fiction led Smith and Frawley to hypothesize that religious discourse has "literary and narrative origins" (p. 369). The researchers also found out that the fourth genre, science, was marked by frequent uses of the hypothetical *if* and the additive *and*. Smith and Frawley explain this arguing that science is characterized by hypotheticality and that the high frequency of additives in this genre is complemented by "the high functional load of the lexicon in this type of discourse" (p. 370). Finally, Smith and Frawley emphasize the essentiality of investigating the types and

distribution of cohesive ties in different discourse types in order to better understand the semantic and rhetorical structure of texts.

Smith and Frawley's findings indicate that genre does have a constraint on the selection of cohesive ties. Such a line of investigation is important in determining the different rhetorical structures of various types of texts. It is also important to studies that attempt to compare compositions written on different topics (see Johnson, 1987). However, more research using large corpuses of texts in a wide variety of genres has to be carried out before any certain conclusions about constraints on cohesive ties can be drawn. Moreover, texts have to be the products of several authors in order to avoid the effect a particular writer's style might have in the selection and distribution of cohesive ties in texts (see Abraham, 1991, below).

In addition to studies of cohesion in written texts, other research has focused on the effect of cohesion on reading comprehension. Irwin (1980) developed two reading passages that dealt with the same subject. One of these, however, had 227 cohesive ties, whereas the other contained 122 ties of Halliday and Hasan's (1976) taxonomy. Irwin then gave these readings to two groups of college students who were told to read them for *understanding*. She recorded their reading times and had them write down all the propositions in the readings that they could recall. Subjects were also instructed to respond to short-answer questions. Results indicated cohesion did affect the number of propositions recalled. Subjects who read the more cohesive prose - as indicated by the higher number of cohesive ties - were able to recall more propositions. Irwin concluded that results of her study supported comprehension theories that emphasize the role of

cohesion in the comprehension process. However, her findings are not supported by other studies to be discussed below, e.g. Chaudron and Richards, 1987, that concluded more cohesive ties are not an indicator of better writing quality.

There are a number of shortcomings in Irwin's study. First, although the two reading versions that she used as material are not provided in the paper, it is clear from her description of these readings that her manipulation of the texts is questionable. She states that the reading version that had a high frequency of cohesive ties, 48 per 100 words, contained more aspects of the 'gibbon' which was the main topic of the passages that she used in her research. This, and not the sheer number of ties, might have had an effect on the subjects' comprehension of the reading passages and their ability to carry out a delayed recall of the propositions expressed in them. Second, since each group of students read one version of the passage, one cannot exclude the reading ability of the subjects as a possible intervening variable that might have affected the results. Third, the fact that the passage containing more cohesive ties had more "argument repetitions", and not the sheer number of cohesive ties, could have been the factor that enabled the readers to do better in the delayed recall test. This seems to be a plausible argument since repetition is a strong factor in enabling readers to transfer information from short-term to long-term memory, and hence aiding them in retaining propositions for a longer period of time.

Likewise, Fishman (1978) carried out a study based on Halliday and Hasan's theory of cohesion. She developed 12 paragraphs with different numbers of noun phrase organizers, or "nouns which introduce and name a category ... and are placed in the initial position in

the paragraph" (p. 160), and references which included personal pronouns, possessives, and demonstratives. These paragraphs were then given to four groups of college students, 119 in all. The subjects read the paragraphs and then answered a multiple choice recognition test. Results indicated there was a sufficient difference in the test scores by noun phrase organizer, but no such differences by number of references. Subjects who read passages that contained noun phrase organizers scored higher than those who read passages lacking these organizers. The researcher, however, argues that these results do not weaken the theory of cohesion because paragraphs containing no references had more repetitions, which might have had an influence on memory, and thus, on subjects' performance on the tests.

Besides examining the effect of cohesion on writing quality and reading comprehension, researchers have also investigated cohesion in spoken English. These studies, however, are relatively few. Planalp, Graham, and Paulson (1987) review studies that attempted to account for cohesion through formal ties analyses. They point out the controversy among researchers as to whether these ties are sufficient to produce cohesion in texts in the absence of conceptual connections. Moreover, Planalp et al. argue the different nature of spoken discourse production warrants the development of a method of cohesion analysis different from the one applied to written texts.

In an attempt to address this question, Planalp et al. (1987) classified formal cohesive ties into syntactic, lexical, and pragmatic devices. Syntactic ties include all the categories in Halliday and Hasan's (1976) taxonomy with the exception of lexical ties. Pragmatic devices, on the other hand, are instances of sequence pairs that are conversationally

associated, e.g. questions/answers, and statements/reactions. The third category, lexical ties, constitutes semantic classes such as synonyms, antonyms, and superordinates.

Planalp et al. then transcribed three different conversations and wrote each turn marked speaker (A) or speaker (B) on a separate card. The cards of each conversation were shuffled and given to a speech communication class at the University of Illinois at Urbana-Champaign. The students were asked to rearrange the cards in the sequence in which they thought the conversation had taken place. Correctly sequenced turns were marked as coherent and incorrectly sequenced ones as incoherent. Planalp et al. then gave the cards consisting of coherent turns to three students from the same group. They were trained in coding cohesive ties into *syntactic*, *pragmatic*, and *lexical* ties. The researchers determined that more lexical ties were present in coherent than incoherent turns. As for the other two categories of cohesion, *syntactic* and *pragmatic*, the difference between their occurrence in coherent and incoherent turns was not statistically significant.

The major flaw of this study concerns the technique involved in its design. And that is the researcher's decision to exclude turns 'incorrectly' regarded to be coherent by the subjects from the analysis of cohesive ties. Since subjects regarded these turns as coherent, there must have been some cohesive traits that prompted them to make that judgment.

Another study that examined nonnative discourse was done by Reynolds (1995) who argues that the investigation of cohesion in second language research has relied on Halliday and Hasan's (1976) taxonomy and Hoey's (1991) concept of repetition. Both approaches, Reynolds claims, regard cohesion as repetition regarding the relation between

two forms. While Halliday and Hasan refer to such a relationship as a tie, Hoey calls it a link. Although the two approaches are similar with respect to the claim that the quantity of cohesive ties is positively correlated to the degree of text cohesion, Reynolds points out, Hoey's approach, unlike Halliday and Hasan's, attempts to account for the significance of individual sentences in a text by the number of bonds a sentence has with the other sentences in that text. However, Reynolds argues, both approaches are based on the false assumption that "quantity equals significance" (P. 189).

To test the hypothesis that quality, and not quantity, determines cohesion, Reynolds (1995) carried out a study in which he investigated cohesion in the written discourse of 26 ESL learners and 16 native speakers of English. The first group was enrolled in an intensive ESL program at the Center for English Language Training at Indiana University whereas the second group consisted of senior students in a composition class at a Georgia high school. Subjects were asked to write essays on which day of the week they liked best and they were required to explain the reasons for their choice. They were given 35 minutes to write the essays.

Reynolds then did a quantitative analysis and a qualitative analysis. For the quantitative analysis he divided each essay into T-units and, using Hoey's (1991) model, compared NS and NNS compositions with respect to: a) repetition frequency, b) lexical repetition/paraphrases, c) link and bond density, and d) occurrence of bonds at paragraph boundaries.

Results of the first measure indicated that both NS and NNS used more simple lexical repetition than any other kind of repetition. However, two-tailed tests showed a

significant difference between the two groups only in the area of simple paraphrases. Results of the second measure indicated that NNS used more repetition than paraphrase; however, a *t* test comparing the means of the two groups did not reveal a significant difference. As for the density of links and bonds, results indicated that both groups were similar in this respect and two-tailed tests did not show any significant differences between NS and NNS. Finally, results showed that although NNS used fewer bonds at paragraph boundaries than NS, the difference was not statistically significant as determined by *t* test.

Concluding that quantitative comparisons of cohesive ties in the writing of NS and NNS showed no significant differences between the two groups, Reynolds conducted a qualitative comparison in which he compared the essays of three NS and three NNS matched by the four measures used in the quantitative comparisons and divided into low, medium, and high usage. The first two essays, the low usage, were very similar with respect to the quantity of cohesive ties. However, Reynolds asserts, the NNS's essay "does not seem as unified or coherent as that of" the native speaker (p. 196). To illustrate his point, Reynolds points out that the NS stated in the introduction that he was going to "write about a weekend day". In so doing, Reynolds argues, "he sets up *day* as the expression of the central topic of the essay" (p. 196). Thus, whenever the word 'day' was repeated it signaled a reiteration of the central topic. On the other hand, the nonnative speaker discussed how people looked forward to her favorite day. However, when she used repetition, she did not do so to achieve her goal. Reynolds quotes the following excerpt to illustrate his point:

[7] I think many people likes Friday. [8] The Friday is end of work day, [9] and They

hope nice weekend. [10] My school finish early than other day, [11] so we can go somewhere we want.

Reynolds argues that one of the problems with this excerpt is that T-unit [9] reiterates the theme of T-unit [7], thus forming an unparallel construction with the preceding T-unit which has 'Friday' as its theme. Another problem is that 'my' in T-unit [10] is "a switch from the generic *people* to the personal *my*" which makes following the writer's argument difficult.

Likewise, similar problems were identified by Reynolds in the essays of the other nonnative speakers with respect to their use of repetition in written discourse. This prompted him to conclude that "a theory of repetition must look not only at overall use of repetition but also at the proficiency of individual uses" (P. 201).

One of the few studies that investigated cohesion in nonnative interview-type spoken English discourse was conducted by Beebe (1980). She developed a taxonomy that consisted of three types of conjunctive adjuncts: 1) simple adverbs, e.g. , *and, but, so, next,* 2) compound adverbs, e.g. *furthermore, nevertheless,* and 3) prepositional expressions, e.g. , *as a result of that.* She applied this model to the analysis of spoken discourse by 19 Asian students from four language backgrounds. The corpus of data was obtained through interviews with a native English-speaking graduate student about the subjects' academic major of study. Results indicated that, contrary to expectations, conjunctive relations were correctly expressed by the ESL learners. Beebe argues, most of the conjunctive adjuncts used by the students were external, and were not intended to convey cohesion. Beebe, however, does not give an example of an external cohesive tie.

She concluded that these adjuncts were redundant because they were used in contexts where they did not contribute to the logical connectedness of the discourse. Furthermore, Beebe could not find a relationship between the proportion of conjunctive adjuncts and subjects' proficiency level or language background.

Another group of studies that dealt with cohesion in nonnative spoken English have investigated academic discourse. This type of research has surged during the past few years as a result of the increase of international teaching assistants (ITAs) whose native language is not English in the American universities. One of these studies was done by Tyler (1992) who investigated the differences in discourse structures and strategies between native and nonnative US English speakers in an academic setting. Tyler argues that previous research has focused on pronunciation and grammatical inaccuracies, hence the need for the investigation of "aspects of the linguistic code which signal logical and prominence relationships" (p. 715).

Tyler used a methodology often employed by researchers in interethnic miscommunication studies to examine the spoken discourse of a Taiwanese graduate student and an American graduate student. The procedure included analyzing the discourse delivered by the speaker, interviewing the speaker to find out his intent about certain parts of the discourse, and gathering information from native English speakers who acted as audience for the discourse. Each of the two graduate students, who had no previous teaching experience, was asked to prepare and give a short introductory lecture about a topic related to his field of specialization. They were further instructed to bear in mind that the audience for the lecture is a group of native English-speaking undergraduate

students who had no experience in the lecturer's area of specialization. The lectures were then videotaped. To control for nonnative pronunciation, the lectures were transcribed and presented by two native speakers of English to a class of 15 native US English speakers majoring in linguistics.

The first group that listened to the Chinese speaker consisted of the ESL instructor, other ITA trainees, and three native speakers of American English. They evaluated the mini-lecture given by the Chinese speaker as roundabout, indirect, and hard to follow. The other group, upon listening to the transcript of the native Chinese speaker's lecture read by a native American English speaker, similarly responded that it was disorderly and aimless. On the other hand, when the transcript of the native English speaker's lecture was read to the same group, none of them described it as disorganized or noncoherent.

Tyler then attempted to account for the differences in perception of the two lectures in terms of lexical specificity, syntax, and interactive effects. The native speaker, Tyler argues, satisfied the conditions of lexical specificity by employing pronominalizations, repetitions, and other lexical choices that made the referent under discussion easily retrievable for the audience. The native Chinese speaker, however, confused his audience by employing non-specific lexical items that made it difficult for them to comprehend his lecture. As for syntax, unlike the native English speaker, the nonnative speaker's discourse consisted of simple clauses. This lack of subordination, Tyler stresses, deprived the audience of essential linguistic forms that "provide cues about prominence, focus, and logical relations" (p. 721). Finally, the researcher partly attributed the difficulties of the nonnative English speaker to what she called "interactive effect". This was exemplified by

the nonnative speaker's use of the anaphoric demonstrative *that* preceding a noun that has not been introduced in previous discourse. This made it difficult for the target audience to comprehend the information.

One of the shortcomings of Tyler's study is that generalizations about the findings cannot be made from investigating the lecture discourse of one nonnative speaker who did not have previous teaching experience. Second, since the conclusion that the usage of nonspecific lexical items and lack of subordination made the lecture of the Chinese subject roundabout and hard to follow was drawn by Tyler and not the audience, it is not evident that they are the real or the only reasons for the negative evaluation the mini-lecture received.

Besides its application to the various areas discussed above, Halliday and Hasan's taxonomy has been used in literary analysis. Karon (1993) states cohesion analysis is common among stylisticians; however, he points out the disagreement as to what constitutes cohesion. He attributes this lack of consensus partly to the confusion of *cohesion* and *coherence*, and to the fact that researchers using Halliday and Hasan's (1976) taxonomy have used only the categories that best suit their varying analyses purposes. Karon commends Halliday and Hasan's model claiming it "provides a wealth of tools of examining texts" (p. 92). However, he argues that interclausal and intersentential ties, like conjunctions, "are syntactic rather than semantic, and thus are not cases of cohesion" (p. 92). Halliday and Hasan's discussion of these intrasentential devices does not, in fact, exclude them from cohesive ties. Their argument is that it is more important to examine intersentential cohesive devices "because they are the ONLY source of texture,

whereas within the sentence there are the structural relations as well ... cohesion is not, strictly speaking, a relation above the sentence" (Halliday and Hasan, 1976, p. 9).

Moreover, Karon states Halliday and Hasan's model is not by itself sufficient to account for cohesion. Applying the taxonomy to the analysis of Andrew Marvell's poem *To His Coy Mistress*, he argues it fails to capture the cohesion between the following two lines:

Had we but world enough, and time,

This coyness, lady, were no crime (in Karon, 1993, p. 93).

Karon states that although Halliday and Hasan's model accounts for some cohesive ties in these two lines e.g., the personal pronouns *we* and *his* (in the title), the conjunctive *but*, and the demonstrative *this*, it fails to explain the relationship implied by the conditional. He concludes cohesion analysis is the analysis of cohesive devices "in terms of conditional connectedness and hence in terms of logic" (p. 102).

The variety of these studies which have applied Halliday and Hasan's model to the investigation of cohesion in a wide range of disciplines indicates that the taxonomy is a useful tool for cohesion analysis. The consensus among the researchers who have applied this model of analysis seems to be that Halliday and Hasan's model is a necessary but not sufficient model for investigating cohesion. Moreover, these studies seem to support Planalp et al.'s claim that researchers modified the taxonomy by applying only the aspects that suit their research goals. Those investigating cohesion in written texts (Witte and Faigley, 1981; and Johnson, 1987), for instance, applied all aspects of Halliday and Hasan's taxonomy to the analysis. This can be attributed to the nature of the model that

lends itself to written texts rather than spoken discourse. The results of these studies, however, were different. While Witte and Faigley determined that more cohesive ties indicated a better writing quality, Johnson concluded that good English essays had more syntactic ties than weak English essays. Semantic ties, on the other hand, did not seem to be an indicator of writing quality.

Conflicting conclusions were also drawn by research that investigated the effect of cohesive ties on reading comprehension. Irwin (1980), for example, determined that more ties facilitate reading comprehension. However, Fishman (1978) concluded that the sheer number of cohesive ties in a text does not affect reading comprehension. She added that noun phrase organizers, or noun phrases that denote the topics and subtopics of the reading passage, do facilitate reading comprehension. It seems that the results of these studies were affected by other intervening variables that resulted from text manipulation. Irwin, for instance, used more cohesive ties together with more repetitions in one of the passages, while she included fewer ties with less reiteration in the other. When the subjects who read the version that had more ties did better in the delayed recall test than those who read the other version, she attributed their better performance to the number of cohesive ties. It is evident that repetition could have had an effect on proposition retention since memorization is very much affected by that factor.

Research that investigated cohesion in spoken language has also applied different aspects of the taxonomy and reached different conclusions. Planalp et al. (1987), for instance, determined that lexical cohesive ties play an important role in defining cohesion while syntactic and pragmatic cohesive ties were not significant in that regard as

determined by statistical procedures. These findings, however, are far from conclusive since Planalp et al.'s procedure of determining cohesive turns in collaborative discourse was based on their subjects' ability to correctly identify consecutive turns by rearranging a shuffled set of cards containing the transcription of the discourse. On the other hand, research that has investigated academic discourse, as exemplified by Tyler's study (1992), has focused on a specific set of ties that denoted logical relationships between segments of a lecture and marked the salience of its significant parts. The findings, however, cannot be generalized since Tyler investigated the discourse of only two subjects, a native and a nonnative speaker.

These conflicting findings seem to be the result of the different taxonomies employed in these studies in addition to the various discourse genres that were investigated. Academic discourse research, for instance, looked into a specific set of markers that are common in this type of discourse (Tyler, 1992). These discourse markers, especially macromarkers, are significant in this genre because academic discourse is characterized by lengthy chunks of discourse that are loaded with information. Since most of this information is 'new' to the audience, these macromarkers seem to be vital in facilitating comprehension by helping the audience organize and understand this type of discourse.

Conversational genres (Planalp et al. , 1987), on the other hand, are more collaborative and are characterized by shorter chunks of discourse. This seems to require a different set of markers such as those identified by Schiffrin and applied by Planalp et al.

As for Planalp et al's. finding that lexical cohesive ties are "found significantly more often" in coherent conversation while syntactic and pragmatic ties are not, it is possible

that this was the result of the research design (p. 339). Since 'cohesion' was determined by the subjects' ability to correctly rearrange the shuffled set of cards containing the transcription of the conversation, it is likely that the subjects relied on discourse topics and subtopics - as marked by lexical items - to figure out the correct sequence of turns in the conversation.

In addition to the research employing Halliday and Hasan's model, other studies have focused on pragmatic cohesion achieved by discourse markers. One of these studies was carried out by Chaudron and Richards (1986) who argue the increasing number of nonnative English speaking students in higher institutions of education in the US makes it necessary to investigate the different aspects of second and foreign language performance in academic settings. The researchers identified three different styles of lectures based on Dudley-Evans and John's categorization (1981) : reading, conversational, and rhetorical style. Then they employed the reading style to examine how the rhetorical organization of lectures as signaled by discourse markers affects the listening comprehension of nonnative listeners. The researchers distinguished between two types of these markers: micro discourse markers e.g., *then, because, but,* and *you see,* and macro discourse markers or "metastatements" that denote key propositions and important junctures in the lecture, e.g., *what I'm going to talk about today* and *let's go back to the beginning.*

The researchers then prepared a lecture on American history and had four different versions of it tape recorded by an ESL teacher. The first version was void of discourse markers and sentence connectors other than those extremely essential to convey the meaning. The second consisted of micro discourse markers and the third of macro

discourse markers. The fourth and last version had a combination of both micro and macro discourse markers. These four versions were then delivered to two groups of ESL students: a pre-university group that consisted of 71 students at a private college in Hawaii and 81 ESL students at the University of Hawaii. To test the subjects' comprehension of the lectures, they were given a cloze test covering a portion of the lecture, a ten-question multiple choice quiz that covered the whole lecture, and a true-false quiz that contained ten questions on the whole lecture.

Results indicated the third version of the lecture consisting of higher-order macro discourse markers "signaling major transitions and emphasis in the lecture" aided the students' comprehension and recall more than micro markers, or "lower-order markers of segmentation and intersentential connectors" (p. 122). Chaudron and Richards concluded macro markers help listeners organize the important points in the lecture as signaled by the lecturer's emphasis. This, they emphasize, helps students build up the proper schema which, in turn, enhances their understanding and compensates for any deficiency they might have in interpreting other aspects of the "rhetorical structure of expository speech" (p. 123).

Chaudron and Richard's study is similar to Tyler's (1992) in that both investigated the distribution of certain linguistic forms that index prominent junctures in academic discourse. However, the design of Chaudron and Richard's study seems to be better than Tyler's for a number of reasons. First, they distinguished between micro- and macromarkers. This distinction is a vital one because, as Chaudron and Richards argued, while micromarkers connect smaller chunks of discourse, macromarkers signal key

propositions and denote important parts of the discourse. Such a finding is important because it could be incorporated in ITA training programs to help nonnative speakers make their academic discourse more effective by employing these important devices in their lectures. Second, they used a larger group of subjects and used a better criterion to test their subjects' comprehension of the lectures by including a cloze test. Such an integrative test is believed to be more valid than discrete point tests.

A similar study was conducted by Williams (1992) who examined the perception of nonnative international teaching assistants (NNITAs) and native-speaking teaching assistants (NSTAs) discourse by native speakers. She reiterated Chaudron and Richard's (1986) concept of macromarkers as expressions that mark important junctures in lectures and used it as a measure of discourse marking. William's study addressed whether explicit marking of discourse is more evident in planned than unplanned discourse. In addition, she also investigated the difference in the degree of text marking between NNITAs and NSTAs in both planned and unplanned discourse.

The subjects for her study were 24 native Mandarin- and Korean-speaking teaching assistants with different majors at an American university. Williams also included another group consisting of 5 NSTAs. She videotaped each TA twice. The first time participants were asked to prepare a 7 to 8 minute presentation on a topic of their choice. They were given a week to prepare their talk and were instructed that no reading would be allowed, however, they were permitted to use lecture notes. The second time subjects were asked to prepare a list of ten introductory topics in their major. One of these was chosen by the instructor and each subject was given 3 minutes to plan the lecture. On the other hand,

the planned discourse of NSTAs was a part of an actual lecture. The unplanned discourse was obtained through the arrangement of the researcher with one of the students to ask the lecturer to review previously presented material.

Videotapes of the lectures were then randomly shown to 25 undergraduates and 10 ESL instructors. They were asked to rate the TA's clarity of explanation and language competence on a 0-to-3 scale similar to the one employed by the Speaking Proficiency English Assessment Kit (SPEAK). In addition to rating the videotapes, viewers were also asked to identify the topic and the main idea of each mini-lecture.

Results indicated NNITAs used more macromarkers in planned than unplanned discourse. Using Chi-squire Williams determined that the differences were statistically significant. No statistical tests were applied to the evaluation of the differences between NSTAs' planned and unplanned discourse because "data sets were not comparable" (p. 702). However, there was only a slight difference in the percentage of markers between planned and unplanned discourse, 65.33% and 62.96% respectively.

Williams then discussed the response of the audience to the different lectures. She stated NSITAs were evaluated by both undergraduates and ESL instructors as more understandable and better at explaining material than NNITAs. Besides, she determined a statistically significant difference between the evaluation of planned and unplanned discourse of NNITAs, the former were given more favorable ratings. Williams concluded that planned discourse of NNITAs was more comprehensible than their unplanned discourse because it employed more macromarkers. This, she added, was not evident in NSTAs. Although they used more markers in planned than unplanned discourse, both

were favorably rated by the audience. Williams attributed this to the NSTAs' ability "to exploit other means of expressing themselves clearly, rather than make extensive use of macromarkers" (p. 706).

Williams used a relatively large number of subjects which makes her results more generalizable. She also tried to account for differences between planned and unplanned discourse of native and nonnative English speaking TAs. However, her usage of a three-point scale to evaluate the TA's performance should have been substituted by a more objective criterion that actually measures the subjects' comprehension of the lectures, e.g. such as the measures used by Chaudron and Richards (1987). Second, when the NSTA's two versions were rated favorably despite the fact that their planned discourse contained more macromarkers than their unplanned discourse, Williams attributed this to the NSTAs' ability to use other expressions to compensate for the lack of macromarkers in their unplanned discourse. On the other hand, when the NNTAs did the same, she argued that the low ratings they received for their unplanned discourse were the result of the fewer macromarkers they used.

In addition to the research investigating discourse markers of non-native speakers in academic settings, there are other studies that have looked into the distribution of these devices in the discourse of native speakers. In a recent study (Elhindi, 1990) I applied Schiffrin's model of local contexts to the classification and analysis of discourse markers that occurred in a telephone conversation tape-recorded from a radio program. The conversation was a dialogue between two adult males, the program presenter and a client.

The program is called Talk Hour and it was recorded during the third week of September, 1990.

The analysis yielded six discourse markers in the data: *y'know, and, well, but, oh,* and *now*. These markers occurred 27 times. Most occurrences were of the marker *well* (nine times), followed by *but* (seven times). *Well* also occurred as an index to three different local contexts. This supported Schiffrin's claim (1985) that *well* is a frequent, multi-faceted discourse marker. Of the twenty-seven occurrences, twenty were indices to the ideational structure. This was partly due to the fact that only the primary function of each discourse marker was analyzed despite the possibility of multiple indexing, i.e. , a discourse marker pointing to elements in more than one local context at the same time. Nevertheless, the twenty occurrences that index the ideational structure conform to Schiffrin's categorization which identifies ideational structure as the local context to which half of the discourse markers that she studied serve as primary indices. The local discourse of ideational structure is a configuration of cohesive relations, topic relations, and functional relations. This huge semantic spectrum also explains the importance of local context.

The results of the analysis support Schiffrin's claim that discourse markers are essential clues to local contexts (Schiffrin, 1987a, 1987b). Speakers use these markers as cues to the expression of their communicative intentions. This can be achieved, as the analyses illustrated, through indicating a shift in the orientation, requiring more information, signaling non-compliance, connecting subsequent utterances to previous discourse, or pointing to a gap in the addressee's knowledge. The analyses also duplicated Schiffrin's

findings (1985a) that *well* is a versatile discourse marker that can be employed to index a variety of local contexts.

Another study that investigated the distribution of discourse markers was that of Abraham (1991) who examined the distribution of the two causal links *because* and *because of* across three discourse genres: mystery fiction, biography, and learned/scientific writing. The genres were represented by an autobiography, two mystery novels, and a book of scientific readings supplemented by three scientific papers. She argues previous studies investigating causal links were restricted to *because* and *so* (Altenberg, 1984; Schiffrin, 1985b). These studies concluded choosing one of these links, rather than the other, is determined by how the causal relationship is sequenced. When *because* is used, the result is mentioned before the cause while the reverse is true for *so*. Abraham argues since the data in Altenberg's study consisted of many more occurrences of *because* (439) than *because of* (34), it is unlikely these two clausal links are interchangeable. She set out to examine the communicative constraints that governed their distribution.

Abraham (1991) explained the syntactic properties of *because* and *because of* as illustrated by:

A. He has spent most of his life in Egypt because he has always been obsessed with finding Nefertiti's tomb.

B. He has spent most of his life in Egypt because of his obsession (p. 324).

In (A) the subordinative conjunction *because* states the cause in an elaborate manner giving detailed information that stretches over a whole clause. On the other hand, *because*

of in example (B) compresses the causality into a short nominalized form. These two causal links offer "alternative structural options that are differentiated by how much detail and elaboration they permit" (p. 325). Although this may appear to be the result of topicalization, Abraham asserts, it is more evident that the distribution of these causal links is governed by how speakers/writers compose their discourse in terms of new/given information. She hypothesizes since new information requires more elaboration, it would be expressed in an elaborate syntactic structure following *because*. However, given information, or information previously known to the listener/reader, wouldn't require such detail and would, more likely, be expressed in a more precise form after *because of*.

To test this hypothesis, Abraham investigated the distribution of *because* and *because of* in three different discourse genres: biography, mystery fiction, and learned/scientific writing. Her analysis yielded 66% occurrences of *because* and 34% occurrences of *because of*. This disparity between the frequencies of occurrence of the two markers, she points out, was consistent with similar studies. Abraham then investigated the constraint of topicality on the distribution of the causal links. To determine what constitutes a topic, she used Brown and Yule's definition of topic as a configuration of "elements derivable from the physical context and from the discourse domain of any discourse fragment" (Brown and Yule, 1987, p. 79). She argued if topicality is a distribution constraint, structures following *because of* would be topical, whereas elaborate structures after *because* would be topic changing. Her analysis, however, did not support this hypothesis.

The second constraint that Abraham investigated was the management of given/new information. She defined *new* information as "situationally given information", or

information that refers to participants or other significant components of the discourse setting, "inferably given information", or information that can be inferred from shared background knowledge, and "contextually given information", or information present in the discourse at hand.

The results confirmed that in fiction and biography *because* encoded new information while *because of* encoded given information. Using Chi-squire, she determined that the difference was statistically significant ($p < 0.5$). As for the deviant genre, learned/scientific fiction, the hypothesis worked in the opposite direction for *because of*. More new than given information was encoded by this marker, 21 versus 6, respectively. Abraham offered two explanations for this: the characteristics of this particular type of writing - this, she argues, was confirmed by Backlund - and "the stylistic idiosyncrasy of the author" (p. 333) whose text she used in her study. Abraham concluded the distribution of these two markers is by no means random as previously believed, and it is not topicality that constraints their distribution, but the management of given/new information.

Abraham's line of research is important because it investigates genre constraints on the selection of ties. This sheds light on how markers that were previously thought to be synonymous, and thus interchangeable, are in fact governed by discourse requirements such as the management of given/new information. However, since her hypothesis was not confirmed by the learned/scientific writing genre, her findings remain questionable. It seems that part of the problem of this study is that two independent variables are involved - genre constraints and given/new information management - while the researcher was

trying to confirm only one of them, the latter. A similar study - discussed above - that drew more concrete conclusions was carried out by Smith and Frawley (1983) who made no assumptions about the distribution of the *conjunctions* prior to the analysis.

Another study that investigated discourse markers was carried out by Ferrara (1995) who traced the evolution of three different types of *anyway* using semantic, syntactic, and prosodic criteria. She analyzed 167 narratives acquired from sociolinguistic interviews in Texas between 1992-1994. The subjects were of different ages ranging from 18 to 80 years and they belonged to different ethnic groups. Ferrara also supplemented her data with narratives in the first person obtained from literature and psychotherapy. She carried out three types of analyses each investigating a different aspect of *anyway*.

The first part of the analysis identified three variants of *anyway*, each with distinct syntactic, semantic, and prosodic characteristics. These three types are illustrated by the following examples:

1. We didn't rent the apartment because it was too expensive. It was in a bad location anyway.
2. It was ugly but he wanted to rent the apartment anyway.
3. He was at the dealership. He always wanted a jag. I thought I heard a noise. Anyway, he decided to buy one.

Ferrara determined that *anyway* in example (1) is an additive semantically similar to *besides*. And it is marked by a level intonation contour. The second *anyway* exemplified by (2), Ferrara pointed out, is a dismissive semantically equivalent to *nonetheless*. This variant is also characterized by a gentle rise followed by a gentle fall in intonation. Both

occur either at sentence-final or sentence-internal position and are not discourse markers. However, Ferrara argues, *anyway* in the third example is a discourse marker that occurs at sentence-initial position and is characterized by a sharp rise followed by an average drop in intonation.

In the second phase of her research Ferrara historically traced the evolution of a number of discourse markers. *Anyways*, she determined, was used as an adverb in the 16th. Century according to Oxford English Dictionary, and was also used as a discourse marker by Charles Dickens in the 19th. Century. She hypothesized it was a northern England dialectal variant that retained its genitive 's'. Other discourse markers she discussed were *anyhow*, introduced as an adverbial in the 18th. Century and as a discourse marker in the 19th, and an earlier variant *anywise*, which was introduced in the 11th. Century and continued to be used until the middle of the 19th.

The final part of Ferrara's paper looked into whether *anyway* is speaker-triggered, used by the speaker to continue her own narrative, or listener-triggered, employed to regain the floor from the listener. She found out that 74% were of the former type and 26% of the latter. Ferrara concluded the discourse marker *anyway* is used frequently in narratives and it amounts to 89% of all occurrences of *anyway* in this discourse type. Moreover, it is employed to indicate shifts in orientation and connect different levels of discourse.

Ferrara's research supplements that of Schiffrin's (1985, 1987a, 1987b) by identifying another marker that has multiple discourse functions. However, her definition of the third variant on the syntactic criterion of occurring at sentence-initial position faces the same problem encountered by Schiffrin (1987b) when she attempted to define discourse

markers. This was not evident in the examples she provided because they were written examples. However, when an analysis of spoken discourse is attempted, the syntactic features of this marker will not help identify it since the boundaries of a spoken sentence cannot be clearly defined. Hence, one should depend on either the prosodic or the semantic features to identify this discourse marker. Moreover, research that investigates the occurrence of this marker in genres other than narration needs to be done before we can fully understand its distribution in different modes of discourse.

The majority of the research using discourse markers has focused on the distributional limitations of a small number of these devices in different genres. This seems to be dictated by the research purposes of the investigators who set out to examine the different functions of a single marker (Schiffrin, 1990), look into important junctures in lecture discourse (Chaudron & Richards, 1986; Williams, 1992), investigate the significance of causal links selection (Abraham, 1991), or historically trace the development of a single discourse marker (Ferrara, 1995).

Besides, these studies are different with respect to the issues investigated. One group of studies, e.g. Chaudron and Richards (1986) and William (1992), looked into the role of micro- and macromarkers in a specific discourse genre, i.e. , academic discourse, determining that the latter are vital in helping organize important points in lectures. Abraham (1991), on the other hand, investigated the constraints of four genres on the distribution of *because* and *because of* using a completely different type of corpus that included an autobiography, two mystery novels, a scientific book of readings, and scientific papers. As for Schiffrin (1987b), she used a quantitative approach to look into

the distribution of discourse markers in sociolinguistic interviews in an attempt to define the cohesive role of these markers.

The diversity of the perspectives from which discourse markers have been investigated seem to result from the fact that this field of research is still in its infancy - the majority of the studies were done in the 1990s - and there are still a lot of areas that need to be explored. Hence, each researcher used the aspects of the model that suited his purposes. Moreover, since the term *discourse marker* does not define a closed set of linguistic forms that belong to a specific category, researchers had the liberty of including in this category a diverse set of forms ranging from a single word, the interjection *oh* (Schiffrin, 1987b), to a whole sentence, *let's go back to the beginning* (Chaudron & Richards, 1986). The interest in this line of research also seems to be sparked by the functional nature of discourse markers which is consistent with the current trend in linguistics that favors the investigation of linguistic forms in actual use.

The extensive research in the field of cohesion, exemplified by the studies reviewed above, indicates how vital this phenomenon is to the understanding and composition of both written and spoken discourse. This type of research is also a departure from the long tradition of considering the *sentence* as the basic language unit that deserves grammatical investigation and analysis. Identifying and accounting for cohesive links between units of discourse implies the construction of a grammar that transcends the limit of the *sentence*.

In addition to this shift in orientation, current research in discourse markers, exemplified by the studies reviewed above, has drawn a number of conclusions that seem to undermine some of the traditional linguistic beliefs. One of these is Schiffrin's (1985,

1987a, 1987b) finding that a diverse set of linguistic forms can have the same function while a single form can have different functions in discourse. Another finding was that of Chaudron and Richards (1986) concerning the academic discourse of nonnative speakers of English. This line of research is a departure from the old tradition that emphasized pronunciation and grammatical accuracy and ignored macromarkers that index prominent junctures in the discourse and hence facilitate comprehension. Beside investigating native and nonnative discourse, this type of research has also determined that the causal links *because* and *because of*, which have previously been considered synonymous and thus interchangeable, are in fact constrained by discourse genre and discourse management (Abraham, 1991).

However, apart from academic discourse studies, research in this area has mainly focused on written rather than spoken discourse. Moreover, it has also employed, to a large extent, formal cohesive ties - both syntactic and lexical - and ignored pragmatic ones. This is partly due to the long tradition of teaching written composition, and to the fact that cohesion is an elusive concept that does not easily lend itself to formal analysis. I believe it could be argued that all *semantic* phenomena have not had their fair share in linguistic theory and research. In addition to these limitations, research in this area has also ignored nonnative spoken discourse for the same aforementioned reasons. The few studies that have investigated nonnative spoken discourse have focused on specific types of cohesive ties.

Hence, future research should investigate the core function of cohesive ties in different genres. Controlling for genre is essential because of the constraints different genres might

have on the frequency and distribution of various cohesive ties. Moreover, models applied to the analysis of cohesion should be expanded since no single taxonomy seems to be sufficient for an in-depth investigation of cohesion. Halliday and Hasan's taxonomy, for instance, would suffice for looking into grammatical cohesion, but it falls short when lexical and pragmatic cohesion are examined. Sciffrin's model, on the other hand, suits pragmatic analysis but fails to account for other cohesive ties.

Despite the extensive research in discourse cohesion, represented by the studies reviewed above, nonnative discourse has not received its due share of investigation. The majority of the research examining cohesion in discourse has focused on written discourse. The reason for this is twofold. Spoken discourse has been ignored as a result of the long tradition of teaching composition. This was based on the assumption that 'writing' is the most important medium in academic settings. Second, the convenience of using 'written' data seems to be a factor in favoring such a type of discourse. Researchers could use preexisting corpuses or use their own instead of going through the trouble of collecting and transcribing spoken data.

Likewise, cohesion in NNS discourse has been ignored for the same reasons. The few studies that have been conducted seem to be necessitated by the recent surge in the number of international teaching assistants at American universities. It could be argued that nonnative nonacademic discourse is equally, if not more, important than academic discourse because it is more frequently used and by a larger number of people in their encounters with both native and nonnative English speaking people who use it as a lingua franca. Such research is vital because it could have implications for teaching programs

intended to help resolve some of the difficulties that nonnative speakers face in their encounter with native speakers.

Investigating nonnative discourse was emphasized by Bardovi-Harlig and Hartford (1995) who argue that the study of interlanguage syntactic and lexical characteristics is not possible at the sentence level. Hence, it is essential to examine the features of nonnative discourse as it is constructed in authentic situations. Bardovi-Harlig and Hartford assert that such research would not only enable us to better understand grammatical and lexical characteristics of interlanguage, but it would enhance our understanding of nonnative pragmatic competence.

One of the few studies that attempted such an investigation was carried out by Young (1995) who compared the Cambridge First Certificate in English (FCE) oral interviews of 11 intermediate and 12 advanced learners of ESL. He looked into the differences between the two proficiency levels as to: a) the mean length of contribution in terms of T-units and minutes, b) the number of new topics that were initiated by the interviewees and ratified, or maintained by their interviewers, and c) the differences in interviewer discourse in the two proficiency levels.

Results showed that high proficiency interviewees spoke faster than intermediate proficiency interviewees as indicated by the mean number of T-units per minute on all three tasks of the interview. High proficiency speakers also spoke more than intermediate ones. However, no difference was determined between the two groups when ratification of topics introduced by the interviewer was examined. Finally, advanced speakers were found to maintain their discourse topics longer than intermediate speakers. Young

attributed the similarity between the two proficiency groups in topic ratification of interviewer-initiated topics to the nature of the interview which obliges subjects to talk on the topics raised by their interviewer.

Although Young's study determined some important findings pertaining to the rate and volume of contributions and topic ratification in the discourse of speakers from different levels of proficiency, it has a number of shortcomings. One of these is the result of using the FCE interview "with its prescribed tasks and reliance on print materials" which may "obscure discourse differences between learners" (p. 36-37). Hence, Young emphasizes the need for an alternative "to the dyadic NS-NNS interview format". A plausible alternative to this kind of interview is the ACTFL OPI. Comparing the two interviews Young argues that the FCE interviews are "highly structured" whereas the ACTFL interviews are "much more free" (p. 28). He also adds that FCE interviews "favor reliability at the expense of validity" by structuring and standardizing the tasks and allotting a specific amount of time for each phase of the interview (p. 29). The OPI, on the other hand, aims "for valid, fluid, interactive assessment" because it does not script any questions. This gives the interviewer the opportunity to vary her questions to match the proficiency level of the interviewee (p. 29). Young believes that this is an indication of favoring validity over reliability.

Despite it's shortcomings, Young's study is important because it looked into the discourse features of NNSs from different levels of proficiency. Future research should build on Young's findings and employ "a dyadic NS-NNS format" that is better than the FCE. A plausible alternative for the FCE, as suggested by Young is the ACTFL OPI

which allows interviewees to perform to the best of their ability. Moreover, such research should investigate the various features of the discourse of NNSs from different proficiency levels.

This study is an attempt in that direction. It uses the OPI format to look into the frequency of cohesive ties in the discourse of nonnative speakers from different levels of proficiency.

This chapter discussed five different models that have been proposed to the investigation of cohesion in discourse, i.e. Halliday and Hasan (1976), Schiffrin (1985, 1987a, 1987b), Charolles (1983), Givon (1992), and Hoey (1991). The chapter also reviewed and evaluated a number of studies that have investigated discourse cohesion applying some of these models. The following chapter describes the method of this study which uses an elaborate model to look into cohesion in the spoken English discourse of Arabs.

Chapter III

Method

The need for studying cohesive ties to determine whether factors other than the sheer quantity of the number of ties are significant in showing differences in proficiency is important. Equally - if not more - important is the investigation of such forms in conversational styles of nonnative speakers (Young, 1995).

Since cohesion could be achieved through the use of a variety of linguistic forms, it is necessary to use a taxonomy that investigates this concept not only within and between sentences but also between turns. This is of utmost importance in collaborative discourse because it reflects the ability of participants to tie their discourse to that of their interlocutors. The model employed for the analysis of the corpus of this study (see below) is designed to achieve this goal. It consists of Halliday and Hasan's (1976) categories of *reference*, *substitution*, *ellipsis*, and *conjunction*, which are local cohesive ties that work inter- and intrasententially, an expanded lexical cohesion category that is both local, when lexical ties spread within the same turn, and global, when lexical ties are reiterated across turns signaling the continuation of the same topic or subtopic, and a set of pragmatic ties that occur turn-initially signaling global cohesion. These are the 11 *discourse markers* identified by Schiffrin (1987b): *oh*, *well*, *and*, *so*, *now*, *then*, *I mean*, *you know*, *but*, *because*, and *or*.

The purpose of this study is to investigate the frequency of these different cohesive ties in the discourse of NNS from three different proficiency levels. The occurrence of each tie will be counted (see coding criteria below) and statistical measures will be used to determine the significance of the differences between the three groups. Beside this quantitative analysis, a qualitative one will be carried out to determine whether the three levels of proficiency use cohesive ties differently.

3.1. Subjects

The subjects (Ss) for this study were 30 Arab adults, 19 males and 11 females. Seven of them are from Egypt, 6 from Sudan, 3 from Libya, 2 each from Oman and Syria, and one S each from Ghaza Strip, Jordan, Kuwait, Oman, Palestine, Qatar, Saudi Arabia, Tunisia, and Yemen. Ss were chosen from one language background in order to control for differences in the use of cohesive ties that may result from language background and/or culture. A homogeneous group with respect to language and culture would also make it possible to generalize the findings to speakers from the same background. The Ss were between the ages of 19 and 41 years. Ss had different levels of educational backgrounds ranging from high school to graduate school.

Ss were chosen from a small community of Arabs in Stillwater. All of them are either members of the Islamic Society of Stillwater, or the Association of Arab Students, or both. At the beginning of the summer of 1995 when I started collecting the data for this research, I explained my intention of collecting the data to Muslim brothers at the mosque where we frequently met to say our prayers. Those who volunteered to participate were then scheduled to be interviewed. Not every subject who was scheduled showed up so

new Ss had to be scheduled. All in all, about 45 people volunteered to be interviewed but only 35 of them actually showed up. Of these interviews, 5 were excluded, 2 because of the poor recording quality and the other 3 because they performed at a level of proficiency not needed for the purpose of this dissertation.

Ss were individually interviewed by a certified OPI interviewer. All the interviews were done in her office.

3.2. Procedure

The American Council on the Teaching of Foreign Language (ACTFL) Oral Proficiency Interview (OPI) was used for assessing the proficiency of the Ss. The ACTFL OPI was chosen to be used as a tool for evaluation because it is thought to be more authentic than other kinds of oral proficiency tests where examinees respond to a cassette player (Madsen, 1983). The assumption is that the ACTFL OPI is a hybrid between an interview and a conversation, therefore it resembles authentic native-nonnative interactions more closely than any other oral proficiency test (Madsen, 1983). Thirty subjects were used for the study, ten from each of the following proficiency levels: intermediate, advanced, and superior. These broad levels were chosen - rather than ACTFL's hierarchy which divides some levels into sublevels, e.g. advanced high, advanced and intermediate-high, intermediate-mid, and intermediate-low - in order to accentuate the differences in the frequency and distribution of the cohesive ties.

Interviews took between 20-30 minutes, depending on the level of the subject. Interviews were tape recorded using a cassette recorder with an external microphone, then they were rated by me and another rater. Both raters took graduate classes in OPI and

worked as research assistants rating OPIs. Whenever there was a disagreement on the level of the interviewee, the tape was given to the interviewer to break the tie. This happened three times during the course of the evaluation. In all three cases the third rater made an assessment that was consistent with one of the two raters. Therefore, for each interview two raters had to agree on the proficiency level. Interviews were then transcribed using a transcription machine. The transcription conventions are shown in List of Transcription Symbols (p. viii).

3.3. Data Analysis

Two models were applied to the analysis of the corpus of data for this dissertation. The first model is based on Halliday and Hasan's taxonomy of cohesion analysis which includes five categories: *reference*, *substitution*, *ellipsis*, *conjunction*, and *lexical* cohesion. The second model of analysis is based on Schiffrin's (1987b) *discourse markers*.

For the first five types of features, *reference*, *substitution*, *ellipsis*, *conjunction*, and *lexical* cohesion, a stretch of discourse on the same topic was chosen for each S (see Appendix A). For superior speakers, the first significant turn, or consecutive turns, that was not less than 200 words, and that expressed or supported an opinion was selected for analysis. The same criterion was applied for the text selection of advanced level speakers. However, since the text type of this proficiency level is different from the superior level, the minimum number of words required for selection was brought down to 150 words. As for the third level of proficiency, intermediate speakers, a different but related genre type had to be chosen for analysis because expressing and supporting a point of view is not part

of their repertoire. The genre type chosen for this group was comparison and contrast. The minimum number of words for this level of proficiency was 150.

3.3.1 Reference

The first category of cohesive ties that was investigated was *reference*. This included identifying each occurrence of *personal*, *demonstrative*, and *comparative* reference. Personal reference, according to Halliday and Hasan's categorization, includes *personal pronouns*, *possessive determiners*, also known as *possessive adjectives*, and *possessive pronouns*. The second subclass of *personal reference* is *demonstrative reference*. This type includes *demonstrative pronouns*, *adverbial adjuncts*, and the cohesive instances of the *definite article* (see below). The last subcategory of *personal reference* is *comparative reference*. This includes a variety of adjectives and adverbs that indicate general comparison, e.g. , *similar*, *identical*, *similarly*, and *identically*, or particular comparison, e.g. , *comparative adjectives* and *quantifiers*.

Coding these types of cohesion was relatively straightforward except for instances in which an item is repeated . This happened in a variety of contexts that included hesitation, emphasis and self-repair. An instance of hesitation is illustrated by the following excerpt from the interview of S #5 when he was talking about polygamy in his country:

[3:1] .. uh .. it .. it means unstability .. for the woman and for the kids ..

In this example only the second 'it' was counted as an instance of *personal reference*. The same subject also produced an instance of self-repair illustrated by the following excerpt in which he was responding to a question about whether polygamy has any advantages:

[3:2] I don't .. I have no experience (laughs) .. but .. but .. it seems to me that .. sometimes it .. it it work out okay ..

In this example 'I', at the beginning of the excerpt was not counted as an occurrence of *personal reference* because it was part of a construction that has been abandoned by the speaker in an instance of self-repair. On the other hand, repetitions, as in the example below, were regarded as multiple occurrences because repetitions have a discourse function, e.g. , emphasis. The following illustration of repetition was by the same speaker discussing polygamy in Lebanese society:

[3:3] I mean .. it's very rare .. it's very rare ..

As for the *definite article* 'the', only occurrences that denoted reference to an item that "provides a target for the anaphora" have been counted (Halliday & Hasan, 1976, p. 72). This excludes instances in which the referent is identifiable for extralinguistic reasons. An example of this is illustrated by the following quotation by S #18 when he responded to a question about the differences between a president and a king:

[3:4] I think .. since I was born .. you know .. the King has been a king .. forty .. two or three years ago ..

Since no previous mention of King Hussein, the King of Jordan, has been made, the meaning of *the* is not dependent on previous text. Therefore it was not counted as a cohesive tie because it does not have a cohesive function. Another instance of *the* that was excluded is when it is used cataphorically to "refer to a modifying element within the same nominal group as itself" (Ibid, p. 72). An example of this type is the following excerpt from the interview of S #6 commenting on how movies portray America:

[3:5] .. when I watch the movies concerning teen age .. people in America .. I see how .. dating .. is something .. important here ..

The last type of *the* that was excluded is when it is used generically to denote a whole class or a member "considered as a representative of the whole class" (Ibid, P. 71). This is illustrated by the following excerpt by S #2 on politics and religion in the Sudan:

[3:6] .. people envision of politics .. as a different thing .. from religion .. you go to the mosque .. and pray .. and and .. perform your religious duties ..

This leaves the only cohesive function of *the* where a target for the anaphora can be retrieved from the text, e.g. ,

[3:7] .. without OSU .. I think the city will be just .. nothing ..

In this example, S #9 is using *the* as an anaphoric cohesive tie to refer to the city of Stillwater which he mentioned by name in his previous turn.

3.3.2 Substitution

The second category investigated in this dissertation was *substitution* which falls into three categories: *nominal*, *verbal*, and *clausal substitution*. *Nominal substitution* is realized by *one*, *ones*, and *same*. This type of *substitution* takes place when a *noun* or a *noun phrase* is replaced by one of these three forms. An instance of nominal substitution is illustrated by the following excerpt of S #3. After discussing military dictatorships in Africa and talking about Idi Amin of Uganda and Numeiri of Sudan, he proceeded to say:

[3:8] .. and right now we have another one similar to that guy ..

In this example, 'that guy' is an instance of *referential* cohesion where *that* is a *demonstrative* anaphorically pointing back to Numeiri whose name was mentioned earlier

in the same turn. *One*, on the other hand, is an instance of *nominal substitution* and it stands for 'military dictator'.

Instances of *one* that are not cohesive, like cardinal numeral 'one' and indefinite article 'one' were not counted as cohesive ties. An example of the former occurred in the discourse of S #10 when he responded to President Halligan's slogan that the university should be more student oriented:

[3:9] well .. this is .. one of those catchy phrases.

And an example of the latter is illustrated by:

[3:10] 'I vote the young lady tells us a story.'

'I'm afraid I don't know one,' said Alice (Halliday and Hasan, P. 101)

The second type of *substitution* is *verbal substitution*. It is realized by *do*. This type of cohesion occurs when *do* substitutes for the lexical verb as head of the verbal group (Halliday and Hasan, P. 112). An illustration of this type is the following quotation by S #2 when he was discussing the regime in the Sudan:

[3:11] .. they can kill people .. and actually they did that ..

In this example, *did* substitutes for the lexical verb *kill*.

The third and final type of substitution is *clausal substitution* which is realized by *so* or *not*. This type of cohesion is exemplified by the following response of S #6 when she was asked whether movies give an accurate view of the American culture:

[3:12] No .. I don't think so ..

In this example *so* is an instance of *clausal substitution* because it substitutes for the clause 'that movies are an accurate view of the American culture'.

3.3.3 Ellipsis

The third category of cohesion investigated in this dissertation is *ellipsis*. Halliday and Hasan argue that "substitution and ellipsis are very similar to each other" and they add that "ellipsis is simply substitution by zero" (p. 142). This type of cohesion falls into three types: *nominal*, *verbal*, and *clausal ellipsis*. An example of nominal ellipsis is the following discourse of S #17. Discussing the Gulf War, he stated that the US deliberately left Saddam in power after his defeat because they wanted to force the Gulf countries to accept the establishment of American military bases in the region. His interviewer reacted by saying "Ah .. isn't that interesting?", to which he responded:

[3:13] it is _ .. and .. they take in every amount of oil .. the country produce ..

In this example 'it is' is an instance of nominal ellipsis where 'interesting', which is a nominal group element, is not expressed.

The second type of *ellipsis*, *verbal ellipsis*, occurs when an item in a previous verbal group is presupposed in a following one. An illustration of this is the following excerpt by S #1, a medical doctor, discussing the time involved in diagnosing an illness in the US, the UK, and Saudi Arabia. Then he added:

[3:14] .. in Sudan .. don't talk about Sudan (laughs) .. you can not _ .. yeah ..

The operator 'can' in this example is an instance of ellipsis since it presupposes the lexical verb 'talk' which is the head of the verbal group.

The third and last type of *ellipsis* is *clausal ellipsis* which occurs when a whole clause is presupposed. This type of coherence is illustrated by the following example by Halliday and Hasan (1974):

[3:15] What are they doing?

Holding hands.

In this example the subject and the modal part of the verb "can be presupposed from what has gone before" (P. 198).

3.3.4 Conjunction

The fourth category of cohesion analyzed in this study is *conjunction*. Conjunctive forms, Halliday and Hasan argue, are cohesive because they specify "the way in which what is to follow is systematically connected to what has gone before" (P. 227). Conjunctions are classified according to the different semantic relationships they express into four subcategories: *additive*, *adversative*, *causal*, and *temporal*.

An instance of additive conjunction is the following excerpt of S #9 on discussing the issue of the importance of the university to the local community:

[3:16] There was a .. major computer company .. trying to locate whether in

Oklahoma or Utah .. and they selected Utah ..

In this example, *and* is an *additive conjunctive* form that simply adds the information 'selecting Utah' to a previous piece of information 'a computer company considering locating in either Utah or Oklahoma'.

The second subcategory of conjunction is clausal relation which expresses result, reason, or purpose. An instance of this type of cohesive tie occurred in the same turn as example [3:16] where the speaker discussed the importance of the university to the local community. In the following excerpt he expressed the reason that prompted the computer company to choose Utah:

[3:17] .. and they selected Utah .. because they have better engineering program ..

In this example, because expresses a causal relationship, i.e. , the selection of Utah and why it was chosen over Oklahoma.

The third subcategory of *conjunction* is *adversative conjunction* which means a relation that is "contrary to expectation". An example of this type is illustrated by the following response of S #2 when he was discussing the issue of religion and politics in the Sudan:

[3:18] .. most people are very religious in Sudan .. however .. they .. they do not particularly .. like to have .. religion .. being **imposed** on them ..

However, in this excerpt, is followed by a proposition that contradicts expectations. One would not expect religious people to refuse religion being imposed on them.

The fourth and final subclass of conjunctive relations is *temporal conjunction* which expresses the sequence of discourse events in time. An exemplification of this type is the following quotation of S #14 on discussing the Israeli-Palestinian conflict:

[3:19] .. there is a peace treaty now .. and Israel .. like the settlement .. they said before that .. Israel support to freeze the settlement there .. but Israel didn't do anything for that ..

The *temporal conjuctive* 'before' denotes that Israel's claim that it would support freezing the settlements was prior to the peace treaty.

3.3.5 Lexical Cohesion

The fifth category of cohesion investigated in this dissertation is *lexical cohesion*. Halliday and Hasan's taxonomy classifies this category into two major types: *reiteration*

and *collocation*. In addition, they subclassify *reiteration* into *repetitions*, *synonyms*, *near synonymy*, and *superordinates*. This taxonomy, however, does not account for many of the semantic or sense relations that have been identified in the literature. Therefore, an expanded model of Halliday and Hasan's was applied to the investigation of lexical cohesion in the corpus of this study. This expanded version keeps the two major categories of *reiteration* and *collocation*; however, it expands the former category to include *paronyms*, *antonyms*, *complementaries*, *directionals*, and *converses* (Lyons, 1977; Cruse, 1986).

The examples below, that are taken from the corpus, illustrate the different types of sense relations that constitute the taxonomy employed in this study. The first type of lexical cohesion is the repetition of the same lexical item. It is illustrated by the following quotation by S #7 when he was discussing the differences between American and Kuwaiti cultures with respect to dating:

[3:20] .. people maybe looking at dating .. as a purely sex .. okay .. people maybe looking at dating in different way .. you know .. I don't know how you define dating ..

This is an instance of repetition where the lexical item 'dating' is repeated twice.

The second type of lexical cohesion is *synonymy*. It is illustrated by the following excerpt in which S #8 is talking about the Libyan regime:

[3: 21] .. it's not even .. you know .. advocating intellectual ideas or .. concepts or anything ..

In this example, 'ideas' and 'concepts' are an instance of *synonymy*.

Near synonymy is the third type of lexical cohesion. An example of this is the following quotation from the interview with S #6. Discussing American movies and how they give the wrong picture of the American society, she commented:

[3:22] .. the problem is .. everybody thinks that .. **all** of America is like that .. and that there are no morals .. no values .. no nothing ..

'Morals' and 'values' in this example are near synonyms.

The fourth type of lexical cohesion is *hyponymy* which is a sense relation that "holds between a more specific, or subordinate, lexeme and a more general, or superordinate, lexeme, as exemplified by such pairs as 'cow': 'animal', 'rose': 'flower ...'" (Lyons, 1977, P. 291). An example of this type is quoted below. It is taken from the interview of S #6 as she was continuing the discussion of the same issue of the last example:

[3:23] .. when I came here .. I .. I didn't see that .. like in the movies it was .. to be honest it was a horrible picture of America .. it was horrible .. they .. I don't see how the media portrays .. America ..

In this example, *hyponymy* is exemplified by the sense relationship that holds between the *subordinate* 'movies' and the *superordinate* 'media'.

The fifth type of lexical cohesion is *complementary*, which refers to ungradable opposites such as *male*: *female* and *dead*: *alive*. An example from the corpus is the following excerpt from the interview of S #8 discussing the Libyan regime of Qaddafi:

[3:24] so if you are not for .. then you are against them ..

In this context, 'for' and 'against' are *complementaries*, or antonyms that divide a particular ideational field into two sections (Lyons, 1977).

The sixth type of lexical cohesion is *paronymy* which refers to the "relationship between one word and another belonging to a different syntactic category and produced from the first by some process of derivation" (Cruse, p. 130). An example of this lexical relation is the following excerpt from the interview of S # 19 discussing the qualities of a good teacher:

[3:25] .. it's very important for teacher to prepare .. everything he looking for .. second thing the goals .. that he should .. sight the goal .. when he teach anything ..

Antonymy is the seventh type of lexical cohesion. It indicates the relationship between gradable opposites like *wide:narrow* and *big:small*. An illustration of this type of lexical relationship is the following quotation by S # 29 comparing buildings in Oman and the US:

[3:26] .. most of it .. is small .. about two doors .. and kind of not .. not that big buildings ..

Antonymy is illustrated by the relationship that holds between 'small' and 'big'.

The last two types of lexical relationships - for which no examples were found in the corpus - are *directionals* and *converses*. *Directional* denotes the relation between lexical items that indicate opposite directions, e.g. *east:west* and *over:under*. Converses, on the other hand, stands for relational opposites. This type denotes the "relationship between two entities by specifying the direction of one relative to the other along some axis" (Cruse, p. 231). An illustration of this lexical relationship is *above:below* and *before:after*.

Despite the expansion of the lexical cohesion model, it is still evident that a whole range of items with a variety of sense, or meaning, relations have to be coded simply as instances of collocation, or items that usually co-occur as a result of some kind of meaning

relationship. Consequently, this category includes items that enter into a wide range of sense relationships, e.g. *patient/ill*, *student/school*, *religion/mosque*, and *tree/fruit*. Halliday and Hasan rightly refer to this kind of cohesion as "the most problematic part of lexical cohesion" (p. 284). The reason behind this difficulty is that no model of analysis that identifies all lexicosemantic relations has been proposed yet. This is due to the large number of lexical items in the language - unlike other cohesive ties whose members belong to a small group of items in a closed set. And it is also the result of the almost infinite number of collocations, or co-occurrences, into which these lexical items can enter.

3.3.6 Discourse Markers

The sixth and final type of cohesion applied to the analysis of the data in this study is *discourse markers*. Since this cohesive relation is different from those exemplified above, and because Halliday and Hasan's is a comprehensive model of analysis that investigates cohesion between sentences, I decided to investigate discourse markers at turn initial position throughout the whole interview for every subject. This decision is also based on Schiffrin's argument that although her model investigates local cohesion "between adjacent units in discourse, ... it can be expanded to take into account more global dimensions of coherence" (1987b, p. 24). One of these global dimensions is to investigate how a speaker is able to achieve coherence by tying her turn to her interlocutor's.

The 11 discourse markers that Schiffrin identified are: *oh*, *well*, *and*, *so*, *now*, *then*, *I mean*, *you know*, *but*, *because*, and *or*. Because these markers were only counted when they occurred at turn-initial position, markers that were preceded by *yeah* and *uhuh*, for instance, were not counted as discourse markers since these forms are lexical items. An

illustration of such instances is the following excerpt in which the S responds to her interviewer's comment that Turkey may be more open minded about women's rights than the US because they have elected a woman for president:

[3:27] uhuh .. and so did Pakistan.

Although *and* in this example is a discourse marker, it has not been counted as one because it did not fall turn initially.

On the other hand, all discourse markers that occurred at the beginning of a turn, or were preceded by pause fillers and other non-lexical stalling devices, such as *umm* and *uh*, were counted. An exemplification of this is the following quotation by S #20 responding to the interviewer's request for information about the location of a city in Florida:

[3:28] uh .. well .. it's .. I think it's in the Atlantic Ocean.

Each one of these categories was investigated in the discourse of the 30 subjects. The first five features - Halliday and Hasan's *reference*, *substitution*, *ellipsis*, and *conjunction*, in addition to the expanded *lexical* cohesion category - were counted in the texts listed in Appendix A using the coding criteria explained above (see Data Analysis). Since the length of these texts varied from one subject to the other, the frequency of each of the five features per 100 words was determined in order to control for the differences in text size. Since these five features mark cohesion by integrating units of discourse within the same turn, the assumption is that they mark local cohesion.

Occurrences of each type were counted and added up and the subcategories of each cohesion type were also identified. In addition, the mean total of each cohesion type was

determined for each of the three proficiency levels and an ANOVA was used to determine the significance of the difference between the means.

As for the last type of cohesion - Schiffrin's 11 *discourse markers* - occurrences were only counted at turn-initial position for each subject. This was based on the assumption that the first five cohesion types were sufficient for a thorough investigation of local cohesion; hence, the need for a measure that evaluates the subjects' abilities to integrate their discourse with their interlocutor's. Since *discourse markers* perform such a function, they were regarded as global cohesive ties. Discourse markers were counted at turn-initial position for each subject throughout the whole interview. Moreover, the mean number of markers per 10 turns was determined for each level to control for differences in turn number. An ANOVA was used to determine the significance of the difference between the means of the three proficiency levels. In addition to computing markers per 10 turns, the ratio of marker per turn for each proficiency level was determined. Although no null hypotheses have been formulated for this study, the expectation was that more proficient speakers would use more cohesive ties in their discourse than less proficient speakers. Hence, the number of cohesive ties was expected to have a positive relation with the level of proficiency.

In addition to the quantitative analysis discussed above, a qualitative analysis was carried out to determine whether proficiency is marked by the quality of the cohesive ties. An example of this type of analysis is illustrated by the following two examples. The first is an excerpt from the interview of intermediate speaker #21 discussing the differences

between the US and the UK; and the second example is a quotation by superior S #6 expressing her views on whether movies are a true reflection of the American culture.

[3:29] All the time very cold .. and rain .. all the time .. and it is .. when I was in London .. it is [extended pause] what I said .. it is too .. small .. than here ... the cars .. you see the cars smaller .. and the .. streets .. when I came here my .. I had my brother in London .. and he will be here before .. in a visit .. and when he .. was .. when I at .. the airport .. he said to me when you go to America .. you find the streets very wide .. and you find the cars very long .. and when I co .. I came here .. yeah .. I saw the streets too .. wide .. and the cars very longs than .. London .. and the people here is friendly than in London.

[3:30] No .. I don't think so .. um .. in certain things yeah .. maybe but uh ... like for instance being a teenager .. when I watch the movies concerning teen age .. people in America .. yeah .. um .. I see other how .. dating .. is something .. important here .. um .. peer .. peers .. groups in .. on campus .. or in schools or in .. that I understand but .. there are certain things that .. when I came here .. I .. I didn't see that .. like in the movies it was .. to be honest it was a horrible picture of America .. it was horrible .. they .. I don't .. I don't see how the media portrays .. America as a .. good place it's .. yeah .. it's the dream of everybody .. in .. uh .. in the Emirates for instance .. or in any Arab country .. oh .. going to America is like .. wow .. the land of opportunity .. um .. but the picture we see is horrible .. my dad .. he came with me .. um .. he stayed for a

week just to check everything out .. because it was the first time for me .. to travel alone ..

The qualitative analysis attempted to explain the way each speaker used conjunctive relations. This included an investigation of the functions of the conjunctions and the different linguistic forms that were used to realize this cohesion type.

This chapter overviewed the models that were applied to the investigation of cohesion in the corpus for this study. These are Halliday and Hasan's (1974) taxonomy and Schiffrin's (1987b) *discourse markers*. The chapter also described how subjects were selected for this study and it explained the procedure employed to interview the subjects and rate their oral proficiency. Finally, the chapter illustrated how the different cohesive ties were categorized.

The following chapter reports the results of both quantitative and qualitative analyses. The chapter also discusses the trends in the use of the different cohesive ties by the three levels of proficiency.

Chapter IV

Results and Discussion

This chapter reports the results of the six types of cohesion that have been investigated in this dissertation. These are *reference*, *substitution*, *ellipsis*, *conjunction*, *lexical* cohesion, and *discourse markers*. The chapter also discusses the findings of the analysis.

4.1 Reference

The first type of cohesive tie investigated in the corpus was *reference* (see Appendix B for typical examples used by each level). Table 1 shows the mean of total reference ties - rounded to the nearest thousandth - per 100 words for each proficiency level and the distribution of the subcategories of this type of cohesion across the three levels.

Table 1

Mean of Reference per 100 Words Across Proficiency Levels

Proficiency Level	Total	Personal Reference	Demonstrative reference	Comparative Reference	SD
Superior	16.300	12.400	3.100	0.900	4.900
Advanced	16.800	12.600	3. 000	1.200	3.600
Intermediate	17.900	12.300	3.100	2.300	2.300

Table 1 shows that intermediate speakers used the highest number of personal references, a total mean of 17.900, followed by advanced and then superior speakers who used a total mean of 16.800 and 16.200 respectively. The highest degree of variation within each group with respect to the use of this type of tie - as indicated by the standard deviation (SD) - was shown by superior speakers at a SD of 4.900, followed by advanced and then intermediate speakers at a SD of 3.600 and 2.300 respectively. The means suggest an unexpected finding since intermediate speakers had the highest mean followed by advanced and then superior speakers.

However, an analysis of variance (ANOVA) determined that the difference between the three levels of proficiency was not significant [$F(2,27) = .491, p > .05$]. These results do not meet the expectation that the number of referential cohesive ties is positively related to the level of proficiency. A similar conclusion was drawn by Bordine (1983) (cited in Reynolds 1995) who discovered that differences in the use of this type of cohesive tie are not indicative of the writing proficiency of NS and NNS. Hence, the results imply that the quantity of references does not mark the quality of written nor spoken discourse. These results also support Reynolds' (1995) claim that research investigating cohesive ties should look into the quality of these ties and how they are used rather than merely count their occurrences. An attempt will be made in the end of this section to carry out such an investigation.

In addition to the findings discussed above, the analysis of the three subcategories of referential ties - *personal*, *demonstrative*, and *comparative* - showed a common trend for all three groups as shown by Table 2.

Table 2

Percentages of Reference Subcategories

Proficiency Level	Total	Personal		Demonstrative		Comparative	
		#	%	#	%	#	%
Superior	393	297	75.33%	75	19.15%	21	5.52%
Advanced	399	301	75.16%	72	17.71%	26	7.13%
Intermediate	325	228	69.07%	55	17.88%	42	13.05%

Table 2 shows that personal reference constituted the highest proportion of the total references for all three levels. Superior speakers used 75.33% personal references, followed by advanced and then intermediate speakers who used 75.16% and 69.07% respectively. This seems to be the result of the high frequency of personal pronouns - which constitute this category - in spoken discourse, specially in an interview genre where the interlocutors make frequent references to themselves and to each other. As for *demonstrative references*, the second most frequently used subcategory by all levels, superior speakers used 19.15%, followed by intermediate and then advanced speakers who used 17.88% and 17.71%, respectively.

Table 2 also shows that the percentage of comparative references used by intermediate speakers -13.05% - was more than double those employed by superior and advanced speakers at 5.52% and 7.13%, respectively. This seems to be the result of topic constraint on the frequency of cohesive ties in discourse. Since the genre of the texts analyzed for

this group of speakers was 'comparison and contrast' (see Appendix A), this may have caused the intermediate speakers to use this relatively high percentage of *comparative* referential ties in their discourse.

Besides these quantitative differences, the analysis also showed qualitative differences in the use of referential *ties* by the three levels. In spite of the large number of *comparatives* used by intermediate level speakers, for instance, a high proportion of these ties was repetition of the same word, specially 'same' and 'different' which amounted to over 50% of the total number of *comparative* referential ties for intermediate speakers, 23 out of 46. Intermediate speaker #27, for instance repeated these two ties 11 times while the total tokens of all *comparative* ties in her discourse was only 12. An excerpt of her interview in which she discussed some of the differences between her home country, Libya, and the United States is cited below.

[4:1] .. we have dates .. like California .. a lot of kind of dates .. and the same flowers .. the same trees .. the same fruit .. (talking about Benghazi, a city in Libya, she eventually added) .. it's different .. Benghazi it is in .. almost in the mountain .. and .. it's different uh .. the weather is a little bit different ..

In this example, the speaker uses 'same' and 'different' 6 times, repeating each one of them twice. The repetition of the same cohesive tie was also evident in the discourse of other intermediate speakers. Subject #23, for example, used a total of 7 *reference* ties, 3 of which were the repetition of 'other' and another 3 the repetition of 'more' while speaker #24 used a total of 4 cohesive ties 3 of which were occurrences of 'different'.

On the other hand, although superior and advanced speakers used a smaller proportion of *comparative* ties, they employed a greater variety of linguistic forms as a realization of this subcategory. Table 3 shows the distribution and type/token ratio of *comparative* ties by the three groups.

Table 3
Distribution and Type/Token Ratio of Comparative Ties

Proficiency Level	Total Occurrences (tokens)	Total Forms (types)	Type/Token Ratio
Superior	21	9	.428
Advanced	25	10	.400
Intermediate	42	9	.214

Table 3 shows that intermediate speakers had the highest occurrences of *comparative* ties in their interviews, a total of 42, followed by advanced and then superior speakers whose total number of this subcategory of tie amounted to 25 and 21 respectively. On the other hand, the type frequency, or number of different linguistic forms that were used to realize this cohesive relation, was very much similar for the three proficiency levels. Advanced speakers used 10 different forms while the remaining two groups used 9 each. Although intermediate speakers used the highest number of comparatives, they used a smaller variety of linguistic forms (types) to realize these comparatives as reflected by the type/token ratio which is 0.214. Superior and advanced level speakers, on the other hand, employed fewer comparatives but used more linguistic forms to realize this cohesive relation. This is indicated by the type/token ratio for these two levels: 0.428 and 0.400 for

superior and advanced speakers respectively. Superior level speaker # 27, for instance, used 4 *comparatives* in his discourse, which is the highest number of this subcategory for this level; however, he used three different forms of ties: 'more', 'better' and 'another'. Therefore, it could be argued that the high proportion of *comparatives* in the discourse of intermediate speakers was the result of both discourse genre that was investigated, i.e. "comparison and contrast', and the fact that these speakers reiterated the same forms of ties as illustrated by their use of 'different' and 'same' which amounted to more than half of the total number of comparatives for this level.

4.2 Substitution and Ellipsis

These two types of cohesive relations will be discussed together for two reasons. The first is that they are quite similar. Halliday and Hasan argue that "ellipsis is simply 'substitution by zero'" (p. 142). And the second reason is that the analysis of these two types of cohesion yielded very similar results as is evident below (see Appendices C for examples of these two types of ties).

Table 4 shows the total number of *substitution* ties and the different subcategories of this type of cohesion in the discourse of the three proficiency levels.

Table 4

Distribution of Substitution Across proficiency Levels

Proficiency Level	Total	Nominal Substitution	Verbal Substitution	Clausal Substitution
Superior	6	3	2	1
Advanced	2	2	0	0
Intermediate	3	1	1	1

As Table 4 shows, very few instances of *substitution* were evident in the corpus. The total number for all three proficiency levels was 11. The highest number for this type of cohesion was used by superior speakers who employed 6 ties, followed by intermediate and then advanced speakers who used a total of 3 and 2 ties respectively.

Similar results were discovered in the analysis of *ellipsis* as shown by Table 5.

Table 5

Distribution of Ellipsis Across Proficiency Levels

Proficiency Level	Total	Nominal Ellipsis	Verbal Ellipsis	Clausal Ellipsis
Superior	3	2	1	0
Advanced	1	1	0	0
Intermediate	0	0	0	0

Table 4 shows that only 4 occurrences of *ellipsis* were counted in the corpus. Three of these were used by superior level interviewees, one by advanced, and none by intermediate level speakers. Because of the small number of *substitution* and *ellipsis* ties, no statistical measures were used to determine the significance of the differences.

The small number of these two types of cohesive relations in the discourse seems to be the result of the nature of these ties and the types of texts investigated in this study. Since *substitution* and *ellipsis* occur when an item presupposes another in close proximity to it in the discourse (see Chapter Three), these types of cohesive relations usually occur in 'adjacency pairs' specially questions and answers. Most of the examples given by Halliday and Hasan to illustrate these cohesive ties were of this kind. Moreover, these two types constituted the smallest number of all types of cohesive relations in the 7 texts that Halliday and Hasan analyzed in the last chapter of Cohesion in English to demonstrate the application of their proposed taxonomy. For instance, the first text they analyzed was an excerpt from Alice in Wonderland in which they identified 19 different cohesive ties. These included every type of cohesion in their taxonomy except *substitution* and *ellipsis*. Of all 7 texts, the only one that included a significant number of these two cohesive relations was a dialogue between an inspector and a lady taken from a dramatic play. The dialogue is characterized by a series of very short turns. Since the texts analyzed in this study consisted of considerably long turns in which the interviewees expressed a point of view or made a comparison, occurrences of *substitution* and *ellipsis* were very few. This also explains why there are no studies - to my knowledge - that exclusively investigated these two types of cohesion.

4.3 Conjunction

The fourth type of cohesive tie investigated in this dissertation is *conjunction* (see Appendix D for typical examples). Table 6 shows the mean of total conjunctions per 100 words for each proficiency level rounded to the nearest thousandth and the distribution of the subcategories of this cohesive relation across the three levels.

Table 6

Mean of Conjunctions per 100 Words Across Proficiency Levels

Proficiency Level	Total	Additives	Adversa-tives	Compara-tives	Temporals	SD
Superior	6.000	4.300	0.800	0.600	0.300	0.800
Advanced	6.900	3.900	1.600	1.000	0.500	1.000
Intermediate	6.800	3.900	1.000	0.800	1.200	2.000

Table 6 shows that advanced speakers used the highest number of conjunctions, a total mean of 6.900, closely followed by the intermediate level at a mean of 6.800 and finally superior speakers whose mean was 6.000. The highest degree of variation in the use of this type of cohesion as indicated by the SD was shown by the intermediate level at a SD of 2.000 followed by the advanced and then the superior level at a SD of 1.000 and 0.800 respectively.

An ANOVA determined that the difference in the use of this cohesive relation between the three proficiency levels was not significant [$F(2,27) = 1.473$, $p > .05$]. However, an interesting trend was found in the category of *temporal* cohesion. Intermediate level

speakers had the highest mean for *temporal cohesion*, 1.200, followed by advanced, and then superior speakers at a mean of 0.500 and 0.300 respectively. A closer examination of the data suggests that this may be the result of individual differences which elevated the mean of the intermediate group. Intermediate speaker # 11, for example, used 'when' 4 times when she was discussing the differences between England and the US.

[4:2] .. when I at .. the airport .. he said to me when you go to America .. you find the streets very wide .. and you find the cars very long .. and when I came here .. yeah .. I saw the streets too wide .. (and in the following turn she added) .. when you go in .. London .. never one .. said to you hi or hello ..

And intermediate speaker # 15 used 3 temporal conjunctions. The 7 *temporal* conjunctions used by these two speakers equal the total number of this cohesive tie used by all superior speakers.

Another possible explanation for the difference in the use of *temporal* cohesion is genre constraint. Because the task of intermediate speakers was to compare and contrast, most of them carried out that task relating past experiences. Since doing so usually requires the use of this cohesive tie to relate events temporally, intermediate speakers employed more instances of this subcategory than superior speakers. All intermediate level speakers had *temporal* ties in their discourse except speaker # 29, who compared eating habits and buildings in Oman, his home country, and the United States without relating any personal experience for illustration.

However, only four superior speakers employed *temporals* in their discourse. This seems to be the result of the genre analyzed for this proficiency level which was expressing

and supporting a point of view. Since most of them did so without the need to relate events on a time axis, six of them did not have any *temporals* in their discourse. The other four superior speakers who used this cohesive tie either related a personal experience to support their point of view or made reference to a number of temporally related events. Superior speaker # 4, for instance, used 5 *temporal* conjunctions when he expressed his view on why the US intervened in the Gulf to end the Iraqi occupation of Kuwait while it did not do likewise in Bosnia.

[4:3] .. remember when Serbia and Croatia .. and Bosnia .. at that time .. they all claimed their independence (eventually he added) .. had the United States stressed at that time .. when Bush was president .. stated that Serbia get out of .. uh .. cut all the aid to the Serbs in Serbia .. and Croatia (then in the following turn he said) .. had these people .. the Bosnian Muslims .. had they been Christians .. oh my God! could you imagine? That's Nixon's word .. before he died.

In this excerpt, the speaker used five temporal ties to make reference to events that happened at different times in his attempt to point out the inconsistencies in American foreign policy.

These differences, though not statistically significant in this study, merit further investigation to determine the extent to which they are indicative of proficiency or of genre. The analysis of the four subcategories of *conjunction* showed distribution trends of these ties that are similar across all proficiency levels. This is evident in Table 7 which shows the percentage of the 4 different types of ties for each level.

Table 7

Raw Number and Percentage of Conjunction Subcategories Across Proficiency Levels

Proficiency level	Total	Additives		Adversatives		Comparatives		Temporals	
		#	%	#	%	#	%	#	%
Superior	147	105	72.18%	21	13.91%	14	9.61%	7	4.30%
Advanced	165	90	56.22%	39	22.50%	23	13.47%	13	7.81%
Intermediate	125	71	56.37%	20	15.67%	14	11.00%	20	16.96%

Table 7 shows that for all levels *additives* constituted the highest percentage of the total number of conjunctions. Superior speakers used 72.18%, followed by intermediate and advanced speakers who used 56.37% and 56.22%, respectively. A likely explanation for this tendency is the high frequency of linguistic forms that realize this subcategory, specially 'and' which is a frequent, multifaceted form. This form, 'and', amounted to 71% of the total additives used by all three proficiency levels, 189 occurrences out of 266.

Another trend indicated by Table 7 is that superior and advanced speakers used *additives* the most, followed by *adversatives*, *comparatives*, and *temporals*. Intermediate speakers, however, employed more *temporals* than *comparatives*. This seems to be the result of relating personal experience which made it necessary for them to refer to a number of temporally related events.

In addition, the three proficiency levels were also different with respect to the variety of conjunctions they used. Table 8 shows the distribution of conjunctive ties and the type/token ratio for each level.

Table 8

Distribution and Type/Token Ratio of Conjunctions Across Proficiency Levels

Proficiency Level	Total Occurrences (tokens)	Total Forms (types)	Type/Token Ratio
Superior	147	15	.102
Advanced	165	16	.096
Intermediate	125	11	.088

Table 8 shows that superior speakers used 15 different conjunctions 147 times; advanced speakers used 16 conjunctions 165 times, and intermediate speakers 11 conjunctions 125 times. As indicated by the type/token ratio, superior speakers used the highest variety of conjunctions at a type/token ratio of .102, followed by advanced then intermediate speakers at .096 and .088 respectively. Not only did intermediate speakers use the smallest number of total forms, but they also used the smallest number of forms for the subcategories *adversative* and *comparative* conjunctions. The only form that they used as a realization of *adversatives* was 'but' which occurred 20 times in their discourse. As for *causals*, they only used 'because' and 'so' which were reiterated 14 times. Superior and advanced speakers, on the other hand, used a relatively wider variety of these conjunctions. Superior speakers employed 4 different *adversatives* 21 times and 3 *causals* 14 times while advanced speakers used 3 *adversatives* 39 times and 3 *casuals* 23

times. The only subcategory in which intermediate speakers had more occurrences than the other two proficiency levels was *temporal* conjunction. They used 5 *temporals* 20 times. Superior speakers used 4 *temporals* 7 times while advanced speakers used 7 *temporals* 13 times. This was perhaps the result of genre as suggested above.

Finally, although the analysis of conjunctions did not show any statistical significance, the above discussion indicates that more proficient speakers use a wider range of conjunctions. It could be argued that the ability to diversify these linguistic forms is one of the factors that make these speakers more proficient. However, the discourse of intermediate speakers, with respect to conjunctions, is more repetitive and monotonous as a result of reiterating the same conjunctions over and over again.

4.4 Lexical Cohesion

The fifth and final type of cohesive tie from Halliday and Hasan's taxonomy that has been applied to the analysis of the interviews is *lexical* cohesion (see Appendix E for examples of lexical ties). Table 9 shows the mean of total lexical ties - rounded to the nearest thousandth - per 100 words for each proficiency level and the SD across the three levels.

Table 9

Mean of Lexical Ties per 100 Words Across Proficiency Levels

Proficiency Level	Total Lexical Ties	SD
Superior	12.200	3.300
Advanced	11.400	2.800
Intermediate	11.400	2.300

Table 9 shows that superior speakers used the highest number of lexical ties in their discourse at a mean of 12.200, followed by intermediate and advanced speakers at 11.400. The highest degree of variation in the use of *lexical* cohesion was that of superior speakers whose SD was 3.334, followed by advanced and then intermediate speakers at a SD of 2.800 and 2.300 respectively. However, an ANOVA indicated that the difference in the use of lexical items between the three groups was not significant [$F(2,27) = .240$, $p > .05$].

A more detailed analysis of *lexical* cohesion showed that a high percentage of the total occurrences of this type of cohesion was either repetition or collocation. Out of a total of 760 lexical ties in the corpus, all subjects used *repetition* 45.53% and *collocation* 31.59%. They only used the other eight subcategories combined 22.88%. These subcategories are: *synonyms*, *near synonyms*, *superoodinates*, *paronyms*, *antonyms*, *complementaries*, *directionals*, and *converses*. Table 10 shows the percentage of *repetition* and *collocation* compared to the other eight subcategories for each proficiency level.

Table 10

Percentage of Collocation and Repetition Across Proficiency Levels

Proficiency Level	Repetition	Collocation	Other Subcategories
Superior	37.50%	36.15%	26.35%
Advanced	54.41%	23.37%	22.22%
Intermediate	45.81%	35.47%	18.72%

Table 10 shows that superior speakers used *repetition* 37.50% and *collocation* 36.15% for a total of 73.65%. All other eight categories combined were used 26.35% by this group. A similar trend is evident in the discourse of advanced speakers who used the highest percentage of *repetition* at 54.41%. This group used *collocation* 23.37% while the occurrences of all other subcategories in their discourse amounted to 22.22%. The trend for intermediate speakers was also similar at a percentage of 45.81% for *repetition* and 35.47% for *collocation*. The percentage of the remaining eight subcategories for this proficiency level was only 18.72%.

Although all proficiency levels were similar with respect to the use of a high percentage of *repetitions* and *collocations*, there is a difference in the distribution of the lexical subcategories. Table 10 shows that superior speakers have a more balanced distribution across the different *lexical* subcategories: 37.50% *repetition*, 36.15 *collocation*, and 26.35% all other subcategories. This implies that superior speakers have a more diverse lexical network than the other two proficiency levels. Advanced and intermediate

speakers, on the other hand, relied heavily on *repetition* which constituted 54.41% and 45.81% for these two levels, respectively. Hence, speakers from these two proficiency levels used fewer instances of the other *lexical* subcategories than superior level speakers.

These results seem to justify Halliday and Hasan's approach which classified *lexical* cohesion into two broad categories: reiteration, which includes *repetition*, and *collocation*. What is implied in this classification is that *repetition* and *collocation* are the most frequent subcategories of *lexical* cohesion in discourse.

The high frequency of these two subcategories seem to be the result of repeating specific lexical items that highlight discourse topic (Brown & Yule, 1983) and the large number of semantic relations that are covered by *collocation*. Since speakers usually repeat key words that denote the discourse topic, instances of repetition are likely to form a considerable proportion of the lexical ties they employ. This is illustrated by the following example from the interview of superior speaker # 1. Discussing the differences between the pace of life in the UK and the US, he was asked what he meant by 'pace of life'. He responded:

[4:4] .. okay .. how fast to do a job .. how fast you can .. solve any .. see your daily problems .. or your .. the daily life .. uh .. for instance .. uh .. here in the US it's very fast .. really fast .. uh ..for instance .. if you'd like to issue a driver license in England .. you need like ten times the time .. which you are going to do it here ..

Since the discourse topic of this excerpt is 'pace of life' and how it is different in the two countries, the speaker gave the example of having a driver license issued to support his

point of view. Consequently, he repeated 'fast' 3 times and 'time' once because both items are important in discussing 'rate' or 'speed'.

Another example to illustrate how the high frequency of *repetition* was a result of reiterating specific lexical items that highlight discourse topic is the following excerpt by advanced speaker # 14. When she was asked to evaluate American policy in handling the Israeli-Palestinian problem, she responded:

[4:5] .. they are doing good job but they .. they need to push .. push little bit harder on Israel .. they are easy on Israel .. they are biased little bit .. well .. I am sorry .. I don't know if it is fair to say they are not .. but the way since I am a Palestinian I feel that .. there is .. uh .. there is a lot of problem Israel support .. there is a peace treaty now .. and Israel .. like the settlement .. they said before that .. Israel support freeze the settlement there .. but Israel didn't do anything for that .. but they are always giving loans for Israel .. money .. everything for Israel .. but they don't push Israel hard for that ..

In this example, repetition is exemplified by 'Israel' (9 times), 'push' (3 times), 'Palestinian' (2 times), and 'settlement' (2 times), all of which are key words in the discussion of the Israeli-Palestinian conflict. These two examples imply that highlighting discourse topic necessitates the repetition of 'key' lexical items that are important for its realization. Hence, the large number of *repetition* in the discourse of all three levels.

On the other hand, the high percentage of *collocation* - the second most frequent *lexical* tie - seems to be the result of the large number of lexical relations that fall under this subcategory. Halliday and Hasan refer to *collocation* as "the most problematic part of

lexical cohesion" and they also call it a "cover term" (p. 284). The difficulty associated with *collocation* seems to be the result of lumping under this subcategory any lexical relation that is not accounted for in the literature. Another difficulty associated with the study of *collocation* is the rather vague definition given to relation. Halliday and Hasan define collocates as "lexical items that regularly co-occur" (p. 284). Since the degree of regularity is hard to define, almost any two lexical items that are semantically related can be coded as collocates. Discussing this issue, Lyons (1977) points out "the impossibility of describing the meaning of collocationally restricted lexemes without taking into account the set of lexemes with which they are syntagmatically connected, whether explicitly in texts or implicitly in the language system" (p. 262). Cruse (1986) uses the term 'collocation' "to refer to sequences of lexical items which habitually co-occur" (p. 40). The only difference between this definition and the one proposed by Halliday and Hasan is that 'habitually' has been used in place of 'regularly'. This, however, does not make the definition clearer since 'habitually' does not determine the degree of frequency of co-occurences.

Another difficulty associated with *collocation* is the lack of a model that defines the various lexicosemantic relations into which collocates enter. This seems to be the result of the fact that lexical items - unlike grammatical forms - do not lend themselves to categorization , a fact which has made semantics the least investigated branch of linguistics. As a result of this, *collocation* became a convenient default category and any lexical relation that didn't fit into other categories was lumped under its umbrella.

This is illustrated by the following example which is an excerpt from the interview of advanced S #15 expressing her views on the OJ Simpson trial.

[4:6] But .. uh .. the thing I .. I don't understand it .. somebody who .. did really commit this murder .. how can he forget .. or how can he be that stupid .. you know .. to go and .. take off his shoes and .. take off his socks and .. throw throw the .. other glove behind his home ..

If we consider the collocates 'commit:murder', 'take off:shoes', 'take off:socks', and 'shoes:socks:gloves', we realize that each set has a different sense relationship. Lumping all these under the subcategory of *collocation* resulted in the large number of *collocates in* the corpus.

Despite the fact that differences in the use of lexical cohesion were not significant, a qualitative look into some of these occurrences in the discourse of the three proficiency levels would reveal that they were used differently. The first example [4:7] is from the interview by superior speaker # 2 and the second one [4:8] is an excerpt by intermediate speaker # 30. The former was discussing the situation in Sudan and the latter was comparing aging, or getting old, in the US and her home country, Egypt.

[4:7] .. a lot of it is .. uh .. relates to the ideological things in Sudan .. you know .. the conceptual things about .. Islam and .. uh .. religious influence .. on the daily life of people .. uh .. most people are very religious in Sudan .. however .. they do not particularly .. like to have .. religion .. being imposed on them .. on their daily life .. they want to pray .. and they want to conduct .. their religious duties .. but they don't want that .. uh .. into their political life .. people envision of politics ..

as a different thing .. from religion .. you go to the mosque .. and pray .. and and .. perform your religious duties ..

[4:8] .. I see here is whole the old people stay alone .. is very bad .. I think so .. yeah .. but my .. my country is uh .. should be the mother or the father stay with the the whole the family .. with .. his son or his daughter or uh .. the people very close over there .. always the .. family uh .. always together and uh .. and talked to the old man and uh .. and see him and uh .. like that .. but I .. here at America is .. whole the old woman and man is alone at .. the big house .. is very .. very .. very scare one .. I said that to my husband .. when we get old .. we should go back .. to my country .. he said yes yes I know that ..

In example [4:7] the superior speaker - among other lexical ties - used the near synonyms 'ideological' and 'conceptual' to explain the orientation of the regime in his home country. He also reiterated 'religious' 4 times and 'religion' twice. Each of these reiterations, however, was used to achieve a different communicative goal. The first time 'religious' occurred in connection with the influence of Islam whereas the second occurrence was to explain the attitude of the majority of people toward religion. The third and fourth times "religious' was used to refer to how people carried out their obligations as Muslims. The speaker also used the lexical chain ' Islam, religion, religious, pray, mosque' to explain how although people in Sudan perform their duties as Muslims, they do not want religion to be imposed on them. On the other hand, the intermediate speaker (example 4:7) reiterated 'old' 4 times and 'alone' and 'people' twice each. Three of these reiterations, however, repeated the same proposition, i.e. , how old people in America are lonely. She

also used the lexical chain 'family, father, son, daughter, husband' but all items in the chain were used to emphasize the same proposition. In sum, although the intermediate speaker had more lexical ties in her discourse, she did not use them as efficiently as the superior speaker did to achieve a number of communicative goals that included explaining the orientation of the Sudanese regime, describing the influence of Islam, and emphasizing the religious disposition of Sudanese people.

Other examples illustrating how the discourse of more advanced interviewees seemed to use lexical relations more efficiently and for a variety of discourse purposes are the following two excerpts from the interview of intermediate speaker # 13 and advanced speaker # 23. Each of these subjects used 25 *lexical* ties. The intermediate speaker used 11 *repetitions*, 2 *paronyms*, 2 *near synonyms*, and 10 *collocates*, while the advanced speaker used 19 *repetitions*, 2 *antonyms*, 2 *collocates*, and 2 *paronyms*. This break down of lexical ties shows how *repetition* and *collocation* were by far the most frequently used lexical subcategories for both speakers. Although these speakers used the same number of ties and subcategories in their discourse, it seems that the advanced speaker did so in a more proficient manner as we shall see below. As mentioned above, the intermediate speaker had 11 instances of repetition. Most of these occurred in the following excerpt in which she discussed the differences in her young son's experience in going to school in the US and her home country, Egypt. Talking about helping her son with his homework in Egypt, she said:

[4:9] .. he need more help and uh .. he needs uh .. a lot of .. a lot of help .. I have to concentrate with him .. all the time .. just now you have to <u>finish</u> <u>quickly</u> .. <u>finish</u>

quickly .. there is another homework .. there's an Arabic homework .. there's a geography homework .. there is a history homework .. it's a lot ..

As the example shows, reiterated lexical items occurred in repeated phrases, e.g. 'finish quickly' - ' finish quickly', or a repeated frame, or structure, e.g. 'there's a ------ homework'. Repeating the same phrase does not add any new information whereas the repetition of the frame ' there's a ------ homework' three times with different school subjects is redundant since the conveyed information was presupposed by 'there is another homework' where 'another' implies any other school subject.

The advanced speaker, on the other hand, used repetition differently as illustrated by the following example in which she discussed what international students should do to fit into the American culture.

[4:10] .. they have to accept the culture where they are in now .. they have to forget about their own culture .. but not their identity (when asked how they could do that, she added) .. I mean .. they have to treat people as they treat them here .. but they have to keep their identity .. they shall be in the melting pot .. I mean .. you know .. the melting pot?

In this example, although the speaker repeated 'culture', 'identity', 'treat', and 'melting pot', each of the repetitions achieves a discourse goal that is different from the first occurrence of the lexical item. The first occurrence of 'culture', for instance, is a reference to the American culture which the speaker argues international students should accept; then she reiterates 'culture' to denote the culture of international students. The reiteration of 'treat' is different from its first occurrence because the first time it was used

to refer to the way international students should treat people whereas the second occurrence denotes the way they are treated by people in the US. Similarly, 'identity' is used the first time as part of her argument that international students should not forget their identity. The second occurrence of this item is a reiteration of the first but this time it is in the affirmative, i.e. , 'international students have to keep their identity', as opposed to the negative 'but not forget their identity'. Finally, the last instance of repetition in this excerpt is a comprehension check. The speaker wanted to make sure that her interlocutor knew what she meant by 'melting pot'.

All in all, more advanced speakers seemed to use reiteration more efficiently and for a variety of discourse purposes. Less proficient speakers, on the other hand, repeated lexical items without adding new information or achieving a communicative goal. Moreover, they reiterated lexical items using the same grammatical structures, or frames, which made their discourse sound repetitive and redundant.

4.5 Discourse Markers

The sixth and final type of cohesive relation investigated in this dissertation is *discourse markers* (see Appendix F for examples of discourse markers). Table 11 shows the mean of *discourse markers* per 10 turns for the three proficiency levels.

Table 11

Mean of Discourse Markers per 10 Turns Across Proficiency Levels

Proficiency Level	Total Markers	SD
Superior	1.979	0.906
Advanced	1.190	0.556
Intermediate	0.912	0.786

This table shows that superior speakers used the highest number of markers in their discourse at a mean of 1.979 followed by advanced and then intermediate speakers who had a mean of 1.190 and 0.912 respectively. An ANOVA determined that the difference was significant [$F(2,27) = 5.266$, $p < .05$]. The highest degree of variation with respect to the use of this type of cohesion was shown by superior speakers as determined by the SD of 0.906., followed by intermediate and then advanced speakers at SD of 0.786 and 0.556, respectively.

The tendency of using more *discourse markers* by higher proficiency speakers was evident throughout the entire interview for all levels. Table 12 shows the total markers and the turn/marker ratio for each level. Turn/marker ratio indicates the number of turns that elapse before a *discourse marker* is used.

Table 12

Total Markers and Turn/Marker Ratio Across Proficiency Levels

Proficiency Level	Total Turns	Total Markers	Turn/Marker Ratio
Superior	828	169	4.89
Advanced	1083	135	8.02
Intermediate	1157	95	12.17

Table 12 shows that superior speakers, despite the fact that they had the smallest number of turns among the three levels, used the largest number of markers, a total of 169. Advanced speakers came second with a total of 135 followed by the intermediate level speakers who used 95 markers. Superior speakers also used markers more frequently than the other two levels. They used a marker every 4.89 turn as indicated by the turn/marker ratio. They were followed by advanced then intermediate level speakers at a turn/marker ratio of 8.02 and 12.17 respectively.

Before discussing these results, let us reiterate the major differences between this type of cohesive relation and the other ones (see Method). First, discourse markers are the only kind of relation that has been exclusively developed for the analysis of spoken discourse (Schiffrin, 1987b). Second, this type of cohesion was the only one investigated throughout entire interviews at turn-initial position whereas other cohesive ties were analyzed in specific excerpts of the interviews (see Appendix A). Closely related to this is the third difference which is the kind of cohesion indicated by *discourse markers*. Because the occurrence of the first five cohesive relations was investigated within and between

units in the same turn, the assumption was that they measured local coherence, or coherence within the same turn. On the other hand, *discourse markers* were counted only at turn-initial position to evaluate the speakers' ability to achieve global cohesion through integrating their discourse to their interlocutors' to achieve not only semantic but pragmatic goals (Schiffrin, 1987b).

Consequently, these results suggest that unlike local cohesive types, global cohesion is a significant indicator of proficiency for this discourse genre. Since the consistent use of these markers entails the ability to achieve the communicative goals indicated by these markers, it could be argued that more proficient speakers are more able to integrate their turns with their interlocutors' turns achieving a number of communicative and pragmatic goals and making the whole collaborative discourse more coherent. Some of these goals are illustrated by the following examples. The first example is an excerpt from the interview of superior speaker # 4 (speaker B) talking about how the Lebanese Civil War forced him to come to the United States.

[4:11] A: You came here because of the war?

B: Oh yes .. oh yes .. I had two choices .. either to be drafted by the militias .. or get out of the country .. basically .. it was very awful at that time ..

Speaker B uses *oh* in this example in a manner consistent with that explained by Schiffrin. Schiffrin states that one of the functions of *oh* in responses to questions is "reorientation to information" when the "questioner may have assumed too much or too little" (p. 86). It is evident in this example that the questioner (speaker A) assumed too little because she probably thought that speaker B fled the country like any other Lebanese who wanted to

evade the war. The information that follows reorients and fills the gap in the questioner's assumption by explaining to her how the speaker in particular was in danger because as a male who belonged to a certain age group he was targeted by the militias as a possible draftee. What the speaker is implying is that it is not only what you assume about the war - that it negatively affected every Lebanese - but also the fact that he was a nineteen year old and all the consequences of being in Lebanon at that age, during that time.

Another example using a different marker is the following excerpt by superior speaker # 1 (speaker B) when he was discussing some statistics about physicians.

[4:12] A: Well .. the statistics say so .. the statistics say that physicians do not have happy families.

B: And they do not have .. happy personal lives .. if you look for instance alcoholism .. among professionals .. you can find that a lot of .. the percentage of alcoholic physicians .. is on the rise ..

Speaker B uses *and* in this example as a "structural coordinator of ideas which has pragmatic effect as a marker of speaker contribution" (Ibid, p.152). Not only that, but the speaker uses *and* in this example to build on his interlocutor's contribution by mentioning another statistical fact that illustrates that not only families of physicians, but also physicians themselves are unhappy. Then he proceeded to support that by stating the rising number of alcoholic physicians indicated by the statistics. This ability to achieve coherence by integrating their participation into the participation of their interlocutors in a cohesive manner was more evident in the discourse of superior speakers than any other level.

Another instance of using a *discourse marker* to achieve global cohesion is illustrated by the following example by superior speaker #10 (speaker B) when he was asked if there were any good leaders in the Middle East.

[4:13] B: .. The closest person that came to be .. a good leader .. was .. Nasser .. in Egypt ..

A: oh wow! .. that's interesting.

B: But .. with all the .. he was .. you could always tell that he meant well .. but his military background .. and the corruption of people .. people swear that he wasn't corrupt but .. but it is kind of hard to believe that .. when you go back and read the unbiased history ..

In this example, the speaker uses *but* to mark "an upcoming unit as a contrasting action" (Schiffrin, 1987b, p. 152). After stating that Nasser came close to being a good leader, the speaker used *but* to signal a unit that contrasts his initial proposition. So what the speaker is saying is that Nasser meant well but he couldn't realize his good intentions because of his military background, the corrupt people around him, and the hint that he himself was corrupt.

The final example of *discourse markers* is the following excerpt from the interview of advanced speaker #18 (speaker B).

[4:14] A: Have you seen any good movies lately? Do you watch movies?

B: I am not interested.

A: Not interested.

B: I mean I don't have time .. for that .. sometimes on TV .. you know ..

In this example speaker B uses *I mean* to modify his previous response to his interlocutor's question. What he is saying in the second turn is that he doesn't go to the movie theater to watch movies not because of the lack of interest as he initially stated, but because of the lack of time. However, he adds, sometimes he watches movies on TV. This is the function of *I mean* that Schiffrin defines as focusing on "the speaker's **own** adjustments in the production of his/her own talk" (p. 203).

In addition to these findings, the results of the analysis of *discourse markers* also demonstrated that *well* was the most frequently used marker in the corpus. Out of a total of 399 markers by all subjects, *well* occurred 168 times. Table 13 shows the occurrences and percentage of this marker across the three proficiency levels.

Table 13

Distribution of 'WELL' Across Proficiency Levels

Proficiency Level	Total	Percentage
Superior	77	45.56%
Advanced	69	51.11%
Intermediate	22	23.16%

Table 13 shows that the trend for using this marker is similar to that of using all other markers, i.e. the number of occurrences increased with proficiency. Superior speakers had the largest number, 77, followed by advanced and then intermediate speakers who had 69 and 22 respectively. The use of this marker by advanced speakers, however, amounted to 51.11% of all markers, followed by superior and then intermediate speakers who used

45.56% and 23.16% respectively. These results bear some resemblance to those discovered by Schiffrin (1985) who argued that *well* is a versatile multifaceted marker that is frequently used in native spoken discourse to achieve a variety of communicative and pragmatic goals.

To illustrate the versatility of this marker, here are some examples from the interviews. The first example is a quote by advanced speaker # 22 (speaker B).

[4:15] A: Can you describe one Arabic dish?

B: Well .. how about 'Kabsa'?

The interviewer's question is grammatically a yes/no question but pragmatically it is a request for information - an instance of an indirect speech act. *Well* which precedes speaker B's response marks his non-compliance with providing the requested information immediately. Instead of describing an Arabic dish, he asked another question - part of an insertion sequence (Brown and Yule, 1983)- to ask for more information before he could respond to the request. When his interlocutor said she was interested in 'Kabsa", he eventually described the dish complying with the request for information.

Another function of *well* is illustrated by the example below which is a quote from the interview of superior speaker # 27 (speaker B), a doctoral candidate in plant pathology.

[4:16] A: .. okay .. in plant pathology .. what kind of employers would that .. who would you be working for?

B: Well .. if you are working in a university situation .. you will be working in .. the extension services .. you would be working in research .. you would be working in water quality .. you would be working in pesticides residues and

things like this .. if you went to the industry .. you will be working with chemical companies .. that manufacture fungicides and .. uh .. agrochemicals .. or you could be working for consulting companies .. you could be working uh .. for government agencies like .. the EPA .. and AFIS and .. things like this.

This quote illustrates a typical function of *well* which Schiffrin (1985) defines as marking a response "whose referent is larger than ... the immediately prior utterance" (p. 657). Since the speaker needed an elaborate response to a seemingly simple question, - the response is larger than what is assumed by the question - he used *well* to mark the discrepancy between his lengthy answer and the short response that his interlocutor probably expected.

The last example of *well* is taken from the interview of superior speaker # 21 (speaker B). After explaining the bad political situation in his home country, the following exchange took place.

[4:17] A: So you are not going back?

B: Well .. not by choice .. it's just basically .. it's kind of .. let's call it exile .. I mean it's a comfortable one .. but it's still an exile of sorts ..

In this example, *well* marks speaker B's divergence from the options given by the yes/no question. Instead of giving a 'yes' or 'no' answer, he explained why he had no choice in making a decision.

Although *well* was the most frequent *discourse marker* used by intermediate speakers as a group - 22 times out of 95 total *markers* - it was only used by 3 speakers from this level. Intermediate speaker #26 used this *marker* 18 times, speaker #29 used it 3 times,

and speaker #24 once. Hence, the majority of intermediate level speakers did not employ this versatile tie. Since *well* is the most frequent and versatile *marker* in the spoken discourse of NSs (Schiffrin, 1987b), it could be argued that the spoken discourse of superior level speakers - and to a lesser extent, advanced speakers - is more native-like than the discourse of intermediate speakers with respect to the use of *well.*

4.6 Summary of Results

These findings indicate that the frequency of *reference* does not mark differences in NNS spoken proficiency. In fact there was a negative relationship between the number of *references* and level of proficiency since intermediate level speakers had the highest occurrences of these ties followed by advanced and then superior speakers. Not only did intermediate speakers have the highest frequency of *references*, but they also employed the highest number of *comparatives* in their discourse. A closer examination of the interviews, however, indicated that the high frequency of this subcategory of *reference* may be the result of genre. Despite the high frequency of *comparatives* by intermediate speakers, they used a small number of linguistic forms over and over again to realize this cohesive tie, e.g. 'different' and 'same' constituted 50% of the total number of *comparatives* in the discourse of this level.

Superior and advanced speakers, on the other hand, used a smaller number of comparatives. However, they employed a greater variety of linguistic forms to realize this subcategory as reflected by the token/type ratio which has a positive correlation with levels. This ratio was 0.428 for superior speakers, 0.400 for advanced, and 0.214 for the intermediate level.

The results of *substitution* and *ellipsis* analysis yielded only 11 and 4 instances of these ties, respectively. The few instances of these two types of ties in the corpus imply that they are not common in discourse marked by long turns, which was the case for the genre investigated in this study. The analysis, however, showed that superior speakers used the highest number of substitutions - 6 occurrences - followed by intermediate and then advanced speakers who had a total of 3 and 2 respectively. The same tendency was evident in the results of *ellipsis* analysis which show that superior speakers had 3 occurrences of this cohesive tie, followed by advanced speakers who used this tie once while intermediate speakers used none. Although these numbers are too small to warrant a statistical test, it may be that superior speakers have the tendency to use more of these cohesive ties than the other two groups.

The results of *conjunction* analysis also indicate that this type of cohesive tie does not significantly mark proficiency as determined. Intermediate speakers, however, used a higher mean of *temporals* than the other two groups. This was attributed to genre constraint on the selection of *conjunctions*. Since the interview task of intermediate speakers was to compare and contrast, and because most of them did so relating personal experience, they employed a larger number of *temporals* than the other two proficiency levels.

Although differences in the use of *conjunctions* by the three levels were also not statistically significant, a type/token ratio suggested that superior speakers used a wider range of linguistic forms to realize this type, at a type/token ratio of .102, followed by

advanced and then intermediate speakers at a type/token ratio of .096 and .088 respectively.

The analysis of lexical cohesion - the last category in Halliday and Hasan's category - were also not statistically significant. A qualitative analysis, however, indicated that superior and advanced speakers used lexical ties more efficiently and for a variety of discourse purposes (see examples 4:7 - 4:10 for an illustration). Intermediate speakers, on the other hand, repeated lexical ties without adding new information. They also reiterated these ties using the same grammatical structure, or frames, which made their discourse sound repetitive and redundant (see example 4:8 for an illustration).

Finally, the analysis of *discourse markers* indicated a significant relationship between the frequency of *discourse markers* and proficiency level. The Superior group, although they had the smallest number of turns of all three levels, used the highest number of *discourse markers* as shown by the turn/marker ratio. They were followed by advanced speakers and then intermediate speakers who had the fewest number of *markers*.

These results indicate that the ability to use global cohesive ties increases with proficiency. This ability is manifested by the use of a variety of *discourse markers* to achieve different communicative goals that include reorientation of information, coordination of ideas, marking contrasting claims, modifying previous response, and marking noncompliance with a request.

The analysis of this cohesive type also yielded a result similar to Schiffrin's (1985a) who discovered that *well* is a frequent and versatile *discourse marker* in the spoken discourse of NSs. This implies that the discourse of superior and advanced speakers is

more native-like than the discourse of intermediate speakers with respect to the use of *well*. Superior and advanced level speakers employed *well* to achieve a number of different communicative goals that included marking deviation from the options given by a question, marking a response whose referent is larger than the previous utterance, and indicating noncompliance with a request asking for information.

In sum, these results indicate that there are no statistically significant differences in the use of local cohesive ties between the three proficiency levels investigated in this study. However there is a qualitative difference between the three levels with respect to the variety of linguistic forms employed to realize these cohesive ties and the communicative goals achieved by using them. As for the use of *discourse markers*, which denote global cohesion, the results showed a statistically significant difference between the three levels. This suggests that the ability to use global cohesive ties increases with proficiency. The following chapter reports the conclusions of this study and discusses the implications of these findings.

Chapter V

Conclusions and Implications

The purpose of this study was to investigate whether the frequency of cohesive ties was different in the spoken discourse of speakers from three proficiency levels. Two types of cohesive ties were analyzed: local cohesive ties, or ties that integrate units of talk within the speaker's own turn, and global cohesive ties that integrate the speaker's turn with his interlocutor's. Although no theoretical hypothesis was explicitly formulated, the assumption was that these cohesive ties increased with proficiency.

5.1. Conclusions

The results of *reference* indicated that there was a negative relationship between this type of cohesive tie and the level of proficiency. This indicates that the sheer number of these ties is not indicative of oral proficiency. Another trend indicated by the analysis of this cohesive relation is that *personal references* constituted a high percentage of *reference* ties - compared to the other three subcategories - for all three proficiency levels. This is the result of the high frequency of 'personal pronouns' in the interview genre. The results also showed that intermediate speakers used more than double the *comparatives* used by both advanced and superior speakers. This, however, may be the result of text type constraint since the texts analyzed for intermediate speakers were 'comparison and contrast', while for the other levels the text type was 'expressing/supporting a point of

view'. The qualitative analysis showed that intermediate speakers typically repeated the same forms of comparatives whereas superior and advanced speakers used a variety of linguistic forms to realize this subcategory. This implies that more proficient speakers have a wider repertoire of *comparatives* which enabled them to diversify the linguistic forms they employed to realize this type of cohesive tie. As a result, their discourse was more detailed and varied. Less proficient speakers, on the other hand, repeated the same *comparatives*. Consequently, their discourse sounded repetitive and redundant.

Another conclusion that can be drawn with respect to *substitution* and *ellipsis* is that these two types of cohesion rarely occur in this genre of nonnative spoken discourse. This is the result of both the nature of these ties and the texts analyzed. Since these two cohesive relations usually occur in 'adjacency pairs' and other genres of interactive discourse that are characterized by short turns and because the texts analyzed in this dissertation were relatively long chunks of discourse that either expressed and supported a point of view or made a comparison and contrast, instances of *ellipsis* and *substitution* were very few in the corpus.

The results also determined that *conjunction* does not mark proficiency levels. However, a trend that was consistent across all levels was that the most frequent subcategory was *additives* followed by *adversatives*, *comparatives*, and then *temporals*. The high frequency of *additives* is the result of the large number of occurrences of 'and'. Moreover, the analysis of *conjunction* shows that intermediate level speakers used twice as many *temporals* as the other two proficiency levels. Again, this may result from genre constraints on the selection of this cohesive tie. Since the task of intermediate speakers

was to compare and contrast, and because most of them did so by relating events from their personal experience which made it necessary to temporally sequence incidents, this group used a relatively high percentage of *temporals* in their discourse.

The analysis of *lexical* cohesion indicated that there was no significant difference between the three levels of proficiency with respect to the number of the *lexical* ties. However, the quality of these ties was different since more advanced speakers used a larger variety of these ties and for a number of different communicative goals that included supporting an argument, making reference to a discourse subtopic, and checking for comprehension. Intermediate speakers, on the other hand, used these ties in a repetitive manner without adding new information. Moreover, they reiterated lexical items using the same grammatical structure, or frame, which made their discourse monotonous and boring. The results of *lexical* analysis also showed that *repetitions* and *collocations* were by far the most frequent categories. The high number of these two subcategories appear to be the result of discourse topic. Since speakers from all proficiency levels typically repeated certain lexical items that denoted the topic of their discourse, repetition constituted a substantial proportion of the lexical ties in the corpus. As for collocation, the high frequency is the result of the lack of a comprehensive lexical taxonomy that categorizes the different sense relations lumped under this subcategory.

Superior speakers, however, used more evenly distributed frequencies of the different lexical subcategories. This suggests that they have a more diverse lexical network than the other two proficiency levels. Advanced and intermediate speakers, however, relied more

heavily on *repetition* and *collocation*. This suggests that these proficiency levels lack the semantic knowledge to realize different lexical subcategories.

Discourse markers, on the other hand, was the only cohesive relation for which the difference was statistically significant between the three levels. Not only that, but the mean use of these markers seems to consistently increase with proficiency. Since these markers are global cohesive ties that achieve coherence between - and not within - turns, it could be argued that global cohesion is the only type of cohesion whose frequency increases with oral proficiency.

Beside the difference in the frequency of *discourse markers*, the analysis revealed a difference in the quality of these ties by the three proficiency levels. Superior - and to some extent - advanced speakers used *discourse markers* to achieve a variety of communicative goals that included building on their interlocutors' discourse contribution, integrating their discourse with their interviewers' and adjusting their own talk. Intermediate speakers, however, did not demonstrate the ability to use these ties in the same manner.

The analysis of *well*, the most versatile *discourse marker*, showed that the discourse of both superior and advanced level speakers had instances that realized the same functions discovered by Schiffrin (1985a ; 1987b) in her analysis of NSs' discourse. Since intermediate level speakers employed this marker less often than the other two groups, it could be argued that the discourse of more proficient speakers is more native-like. The majority of intermediate level speakers - 7 out of 10 - had no occurrences of *well* in their discourse.

These findings imply that using a large number of local cohesive ties does not mark proficiency. Although intermediate speakers used as many - and for some types more - cohesive ties as superior and advanced speakers, their discourse was repetitive and monotonous as a result of reiterating the same linguistic forms in the same grammatical structures, or frames, without adding new information. More proficient speakers, on the other hand, were able to use a wide variety of linguistic forms to realize the different local cohesive ties. They also used these ties to achieve different communicative goals. The ability to diversify these ties resulted in more varied and detailed discourse by these speakers.

As for the global ties, the findings imply that more proficient speakers had the pragmatic awareness to realize that cohesion in dyadic discourse is the responsibility of both interlocutors. Moreover, they had the ability to contribute to global discourse cohesion by employing the necessary *discourse markers* that integrate their discourse with their interlocutors'. Since intermediate level speakers did not use *discourse markers* as often, it could be argued that they lack the ability to contribute to the global cohesion of dyadic discourse.

In sum, for local ties it is not the number of cohesive ties that mark proficiency but how these markers are used. As for global ties, both the quantity and quality of these cohesive relations marked the level of proficiency. The results support the finding of Reynolds (1995) who stresses that research investigating cohesion should look into the quality of cohesive ties and how they are used rather than merely count the occurrences of these ties.

5.2. Implications

Before discussing the implications, the limitations of this study should be stated. This study has two limitations. The first one is that Ss were not randomly selected because of the small number of Arabs in Stillwater. The second limitation is that the text type analyzed for intermediate speakers - comparison and contrast - is different from that analyzed for superior and advanced speakers - expressing and supporting an opinion . Since there was no single text type that was present in all interviews, I decided to use comparison/contrast for intermediate speakers because it is a task that often involves expressing an opinion. Because of these limitations, the conclusions of this study are only tentative.

Research investigating nonnative spoken discourse is significant to a number of disciplines that include language teaching and interlanguage development. Such research sheds light on how nonnative speakers from different levels of proficiency use cohesive ties in their encounter with native speakers to achieve their communicative goals. The findings of this study suggest that the sheer quantity of local ties does not mark proficiency. Hence, teachers should focus on how these ties are used rather than introduce whole lists of them as the common practice is in ESL/EFL. In addition, the ability to use *discourse markers* - the only cohesive tie that marks proficiency - in face-to-face conversations should also be taught to NNSs. This would improve the pragmatic ability of ESL/EFL learners and enable them to achieve global coherence in collaborative discourse.

Another discipline that could benefit from research in this area is language testing. Using global cohesive ties could be integrated in the criteria for oral proficiency assessment because they are the only type of cohesion that marks proficiency as determined by the findings of this study.

The findings of this study could also be integrated in the guidelines of oral proficiency interviews like the ACTFL OPI. The present guidelines of this test consist of four areas: global tasks, text types, accuracy, and context. None of these, however, assesses the pragmatic ability of examinees. Since the frequency of *discourse markers* consistently increased with proficiency - as determined by the results - descriptors that assess this ability should be added to the OPI guidelines to evaluate interviewees' ability to achieve global cohesion.

In sum, this study determined that the quantity of local cohesive ties does not mark proficiency. However, more proficient speakers use a wider variety of these ties to achieve different communicative goals. On the other hand, the frequency of *discourse markers* increases with proficiency level which implies that using these *markers* demonstrates the ability to achieve global cohesion in collaborative discourse. These findings suggest that the quality - rather than the quantity - of cohesive ties in discourse marks differences in the oral proficiency of NNSs.

References

Abraham, E. (1991). Why 'because'? The Management of given/new information as a constraint on the selection of causal alternatives. Text, 11, 323-339.

Altenberg, B. (1984). Causal linking in spoken and written English. Studia Linguistica, 38, 20-69.

Bardovi-Harlig, K. & Hatford, B. (1995). The construction of discourse by nonnative speakers. Studies in Second Language Acquisition, 17, 125-128.

Beaman, K. (1984). Coordination and subordination revisited: Syntactic complexity in spoken and written nonnative discourse. In D. Tannen (Ed.), Coherence in spoken and written discourse (pp. 45-80). Norwood, N.J.: Ablex Publishing Corporation.

Beebe, L. M. (1980). Cohesion in interlanguage: A study of conjunction. (Report No. 143). New York: Teacher's College, Columbia University. (ERIC Document Reproduction Service No. 268806)

Brown, G. & Yule, G. (1983). Discourse Analysis. Cambridge: Cambridge University Press.

Carrell, P. L. (1982). Cohesion is not coherence. TESOL Quarterly, 16, 479-489.

Charolles, M. (1983). Coherence as a principle in the interpretation of discourse. Text, 3, 71-97.

Chaudron, C. & Richards, J. (1986). The effect of discourse markers on the comprehension of lectures. Applied Linguistics, 7, 113-127.

Cruse, D. A. (1986). Lexical Semantics. Cambridge: Cambridge University Press.

Derrick-Mescua, M. & Gumca, J. (1985). Concepts of unity and sentence structure in Arabic, Spanish, and Malay. (ERIC Reproduction Service No. 260590)

Dudley-Evans, A. & Johns, T. F. (1981). A team teaching approach to lecture comprehension for overseas students. In The Teaching of Listening Comprehension. London: The British Council.

Elhindi, Y. (1990). Discourse markers in conversational English. Unpublished manuscript.

Ehrlich, S. (1988). Cohesive devices and discourse competence. World Englishes, 7, 111-118.

Fishman, A. S. (1978). The effect of anaphoric references and noun phrase organizers on paragraph comprehension. Journal of Reading Behavior, 10, 159-170.

Ferrara, K. (1995). Social and linguistic variation in discourse markers: The case of *anyway, anyways, anyhow, at any rate.* Paper presented at the meeting of the American Association of Applied Linguistics, Long Beach, CA.

Foster, D. (1984). Coherence, cohesion, and deixis. (ERIC Reproduction Service No. ED 245223)

Givon, T. (1992). The grammar of referential coherence as mental processing instructions. Linguistics, 30, 5-55.

Halliday, M. A. K. & Hasan, R. (1976). Cohesion in English. London: Longman.

Hoey, M. (1991). Patterns of lexis in text. Oxford: Oxford University Press.

Hymes, D. (1974). Foundations in sociolinguistics: An ethnographic approach. Philadelphia: University of Pennsylvania Press.

Irwin, J. W. (1980). The effect of linguistic cohesion on prose comprehension. Journal of Reading Behavior, 7, 324-332.

Johnson, P. (1987). Cohesion and coherence in compositions in Malay and English. (ERIC Reproduction Service No. 341229)

Karon, J. (1993). Cohesion as logic: The possible worlds of Marvell's "To His Coy Mistress". Style, 27, 91-105.

Lyons, J. (1977). Semantics (Vols. 1-2). Cambridge: Cambridge University Press.

Madsen, H. S. (1983). Techniques in testing. New York, NY: Oxford University Press.

Mandler, J. (1978). A Node in the code: The use of a story schema in retrieval. Discourse Processes, 1, 14 - 35.

Planalp, S. ; Graham, M. & Paulson, L. (1987). Cohesive devices in conversations. Communication Monographs, 54, 325-343.

Reynolds, D. W. (1995). Repetition in nonnative speaker writing. Studies in Second Language Acquisition, 17, 185-209.

Schank, R. & Abelson, R. (1977). Scripts, goals, plans and understandings. Hisdale, NJ: Erlbaum.

Schiffrin, D. (1985a). Conversational coherence: The role of "well". Language, 61, 640-667.

(1985b) . Multiple constraints on discourse options: A quantitative analysis of causal sequences. Discourse Processes, 8, 281-303.

(1987a) . Discovering the context of an utterance. Linguistics, 25, 11-32.

(1987b) . Discourse markers. Cambridge: Cambridge University Press.

(1990) . Between text and context: Deixis, anaphora, and the meaning of *then*. Text, 10, 245-270.

Smith, R. N. & Frawley, W. J. (1983) . Conjunctive cohesion in four English genres. Text, 3, 347-374.

Tyler, A. (1992) . Discourse structure and the perception of incoherence in international teaching assistants' spoken Discourse. TESOL Quarterly, 26, 713-729.

Van Dijk, T. & Kintsch, W. (1983) . Strategies of discourse comprehesion. New York: Academic Press.

Williams, J. (1992) . Planning, discourse marking, and the comprehensibility of international teaching assistants. TESOL Quarterly, 26, 693-711.

Witte, S. & Faigley, L. (1981) . Coherence, cohesion, and writing quality. College Composition and Communication, 32, 189-204.

Young, R. (1995) . Conversation styles in language proficiency interviews. Language Learning, 45, 3-42.

APPENDICES

Appendix A

Texts for Local Cohesion Analysis

Level & No.	No. of Words	Topic
Superior 1	223	views on pace of life in the Uk, the US, and Saudi Arabia
Superior 2	234	views on politics and religion in Sudan
Superior 3	226	how dictators get into power in Africa
Superior 4	277	why the US did not intervene in Bosnia
Superior 5	253	views on polygamy in Lebanon
Superior 6	289	whether movies are a true reflection of the American culture
Superior 7	240	what constitutes culture and how it is different in Kuwait
Superior 8	236	what is the Libyan regime like
Superior 9	205	the impact of OSU on the local community
Superior 10	256	differences between graduates and undergraduates
Advanced 1	244	did America do the right thing in the Gulf?
Advanced 2	272	how children should be brought up in a materialistic culture
Advanced 3	331	what international students need to know about the American culture
Advanced 4	277	how should America handle the Middle East Problem
Advanced 5	204	thoughts on the OJ trial

Advanced 6	150	what is the ELI like and how it should be improved
Advanced 7	214	did the US want to take Saddam down during the Gulf War
Advanced 8	210	is a king different from a president
Advanced 9	244	what are the traits of a good teacher
Advanced 10	222	why the interviewee decided to go into business
Intermediate 1	186	comparing England and the US
Intermediate 2	203	differences between Saudi Arabia, Sudan, and the US
Intermediate 3	230	differences in school systems between the US and Egypt
Intermediate 4	203	status of women in Yemen and the US
Intermediate 5	174	differences in experience in two university towns
Intermediate 6	155	differences between Tunisia and the US
Intermediate 7	180	comparing Tripoli and California
Intermediate 8	109	differences in education in Egypt and the US
Intermediate 9	200	differences between Oman and America
Intermediate 10	154	aging in Egypt and the US

Appendix B

Examples of Reference

1. Superior

Uh .. That's the question which raises itself .. you know .. these people actually .. you know .. they are like .. high school dropouts .. that's another .. that's another problematical issue .. in .. in Africa in general .. you know .. those high .. high school dropouts .. they .. they join the military .. and then they start .. putting those .. you know .. stars on their shoulders .. and became .. very much recognized .. okay .. you know .. exchange from a lieutenant to .. another rank .. and .. higher ranks .. and so forth .. and then suddenly .. you feel one of them .. he just .. want to oppress people and became the president of the country ..

2. Advanced

Yeah .. but they told us that America tend to be like a melting pot .. each body comes here .. becomes part of the culture right here .. so some .. some of them forget their own identities .. and this is not good .. this is no good .. they have to keep their own roots and their own identities .. and the .. they have to accept .. the behavior of the people here because .. it's completely different from our owns .. but we have to accept it .. and it's .. I don't mean that it's different it's bad no .. sometimes it's better .. I like lines here by the way (laughs) .. here there are long lines sometimes .. okay .. but I like them.

3. Intermediate

Visi .. I like when I been .. to California .. if you like .. I mean .. in my .. in my city exactly .. Tripoli .. we have different uh .. different weather and different kind .. between city and city .. we've Tripoli the capital .. and Benghazi .. it's the second uh .. city ... uh .. but .. you know .. it's between about .. I don't know exactly with .. between Tripoli and Benghazi about .. one thousand two hundred kilometer .. I don't know exactly in the mileage .. but .. it's different .. Benghazi it is in .. almost in the mountain .. and .. it's different uh .. the weather it is a little bit different .. like Oklahoma and uh .. Houston .. little bit .. but Tripoli Exactly like uh .. like California.

Appendix C

Examples of Substitution and Ellipsis

1. Superior (substitution)

A: You bring up an interesting point about movies .. which I would like to pursue a little bit .. we .. represent our cultures through our movies .. to our own members and to the rest of the world .. and I wonder whether that's .. an accurate view of our culture.

B: No .. I don't think so .. um .. in certain things yeah .. maybe but uh ... like for instance being a teenager .. when I watch the movies concerning teen age .. people in America .. yeah .. um .. I see other how .. dating .. is something .. important here .. um .. peer .. peers .. groups in .. on campus .. or in schools or in .. that I understand but .. there are certain things that .. when I came here .. I .. I didn't see that .. like in the movies it was .. to be honest it was a horrible picture of America ..

2. Advanced (ellipsis)

B: Before the Gulf .. before Iraq .. uh .. invaded Kuwait .. ten years ago .. uh .. America gave uh .. an option to the Gulf states .. to just put one base in any .. in any country of the six countries .. and pay the Gulf countries money .. you know .. as much as they want .. uh .. because of the .. you know .. Iraq-Iran .. this movement .. they are saying that they are scared of .. or .. trying to secrets .. or take some secrets .. and the .. Iraq I mean .. the Gulf countries .. said no

.. you know .. any option the American gave to the .. uh .. Gulf countries .. and the money .. they offer .. the Gulf countries say no .. so .. it was for the .. for their .. I mean .. for the American good .. I mean .. the Iraqi to invade Kuwait .. you know .. to put in every country a base .. and instead of paying their countries .. I mean .. my country and Saudi and Kuwait .. paying the Americans .. okay .. and if they take the Saddam .. down-

A: Then there is no reason for the base?

B: And we will tell them .. you know .. I am telling you .. go out of my country.

A: Ah .. isn't that interesting?

B: It is ---- .. and .. they take in every .. amount of oil .. the country produce .. they take some taxes .. and .. you know .. beside the money they .. we pay them .. so.

3. Intermediate (substitution)

A: Can you you compare it (interviewee's hometown) to Stillwater?

B: Well .. it's um .. the most people are students there .. but they are .. I mean we haven't .. uh .. college there uh .. the other big cities .. the nearest one .. uh .. about .. uh .. eighteen mile ..

Appendix D

Examples of Conjunction

1. Superior

And by the way .. I am a medical doctor .. I worked in Sudan .. and in Saudi Arabia .. and in England .. as a medical doctor .. so .. this is absolutely pertinent to my field .. that speed which I am talking about .. for instance .. in the US .. in like twenty four hours .. you can get all the details you need about any disease .. or about any patient .. and you can .. come up with .. a hundred percent sure diagnosis .. in England you need .. more time .. in Saudi Arabia .. definitely .. in Sudan .. don't talk about Sudan (laughs) .. you can not .. yeah.

2. Advanced

Yeah .. I mean .. uh .. I hear some people .. who were interesting of .. changing for example the .. way of rule from the king to president .. all they like .. the president of our neighbor country ... and they were in uh .. some parties .. which .. maybe against .. but later .. they discovered that .. the king was better than .. others .. you know .. from their own experience .. so I don't know .. for me .. I never practiced anything .. but some people were in parties .. for example .. which are with .. another government .. which is close to our country .. they have different ideas .. different opinions .. but they said that they were feeling .. and they found out the things .. comparing to the others .. at least in his age .. and in the area .. he is better than others.

3. Intermediate

Yeah .. they do.. they don't learn so much .. but even if you are going .. the school are going to give them .. a lot of uh .. information .. they are going to accept it .. because the way the .. the teachers giving them .. the way they are learn .. you know .. in our country .. they study more .. they study more .. because it is a lot .. a lot of things .. but here .. they took the .. these info .. the small information .. and the teacher keep .. going on .. just speak about this information .. you know.

Appendix E

Examples of Lexical Cohesion

1. Superior

A: I don't know much about differences .. what does that mean 'fundamentalist part'?

B: These are people .. who believe that .. they can implement .. the religious rules .. if you will .. to the daily life of people .. they want .. uh .. and they do their own interpretation .. of these rules .. according to what they see fit .. and not necessarily everybody agrees with that .. I certainly don't agree .. and .. uh .. to the point .. they can implement that using .. whatever .. means .. they have available .. they can kill people .. and actually they did that .. they can imprison people .. they can torture people.

2. Advanced

But .. uh .. the thing I .. I don't understand if .. somebody who .. did really commit this murder .. how can he forget .. or how can be.. can he be that stupid .. you know .. to go and .. take off his shoes .. and take off his socks and .. throw the .. other glove behind his home .. he should have been .. uh .. a little bit more .. clever .. to leave everything there .. not to throw them .. at least not to throw them at his house .. throw them .. somewhere else.

3. Intermediate

Yeah .. in that time .. really I improved my language .. higher .. it's not just by education .. no it's .. it's look like .. uh .. speak .. speak .. I tried to fix my grammar .. or I fix my speech .. about .. pronounce something .. it's uh .. different .. and after that uh .. I spent maybe .. three semesters .. or something like this .. then I transferred last summer ..

Appendix F

Examples of Discourse Markers

1. Superior (example 1)

A: What would you like to do?

B: Well .. you know .. I have really a strong record in publication .. I have about thirty- two publications .. and .. in very professional journals .. you know .. in chemical engineering .. environmental engineering .. uh .. I think I have a strong foundation .. my background .. uh .. I would want really to pursue .. you know .. chemical plus you know .. implementing .. you know .. the basic fundamental in chemical engineering .. to .. for solution in environmental part .. and you know .. you know .. this is the second .. you know .. motivation why I go for environmental engineering .. you know .. having the background .. the strong background .. it doesn't mean .. you will involve really in the .. what's going on in the real world .. when you go to environmental engineering .. well .. they have a lot of problems .. okay .. but their background .. most of the faculty there .. their background .. do not really qualify them .. does not really qualify them .. to pursue .. you know .. advanced things you know .. in environmental engineering.

Example 2

A: So you think the government is making them fight.

B: Well .. it's .. it's the government or .. the .. people who control the .. the arms .. people make a lot of money out of war .. millionaires were created in Lebanon ... just .. because of the sales of the arms .. because of the .. uh .. monopoly in food .. monopoly in water .. the monopoly in medical supplies.

2. Advanced (example 1)

B: In nineteen eighty eight .. Gaddafi decided .. the only way to .. get rid of opposition groups outside Libya .. is by .. uh .. uh .... how I can find the word .. okay .. is by breaking .. this organizations .. so he found out the weak points in these organizations .. the weak points is .. most Libyans are not .. good .. uh .. immigrants .. they don't like to live abroad .. they like to go back .. to their .. uh .. country and live with their families and .. you know .. live in the same culture .. and .. he decided that .. many of them has lived .. abroad for long periods of time so .. the best thing for him is .. to give them .. an opportunity to come back home .. so he opened up and he said .. anybody who comes back to the country nothing will happen to him.

A: Could you trust him?

B: And some people did.

A: Yeah .. could you trust him?

Example 2

A: You have any free time?

B: Oh .. this time not .. this time because I am .. almost finishing my Master's degree right now .. and I am doing hard work in my research **but** .. in .. let me

talk to you about the past time .. uh .. in my free time usually .. uh .. I love to gather with my friends .. at the beginning I came here .. we have .. many organization here .. we have many places we go to .. uh .. we have Saudi Student House ... which is .. I was the president of that house .. and .. association .. right now I am the .. financial secretary of that place .. and organization .. I have the .. Islamic Society of Stillwater .. I love .. traveling uh ... uh .. I just came from Chicago .. I spent there two days with my family .. and my children.

3. Intermediate (example 1)

B: Uh .. maybe in my country .. if .. if my husband .. uh ... finished his studies .. we can move whether to .. -

A: But I thought you said .. there is not enough money in your country.

B: But .. my .. my husband is a medical doctor .. when .. when when .. uh .. with this .. uh .. specialization ... you can find good job at any where .. and after that .. in my country if you have money .. you can live.

Example 2

G: So you suffered because of the war.

E: Yeah .. dollar start to go up ... and ... everything start to change .. Dollar was like uh .. ten riyals .. Yemeni .. for our money and now .. dollar's cost two hundred riyal Yemeni.

G: Wow!

E: So you can see the big difference .. people couldn't find job .. uh .. everything is there .. food .. vegetable .. everything .. but .. you need the money ... government

doesn't pay really good salaries ... they .. they give like ... something very .. you you can not live with our salary here .. never.

2

VITA

Yousif Elhindi

Candidate for the Degree of

Doctor of Philosophy

Thesis: COHESION IN THE SPOKEN ENGLISH DISCOURSE OF ARABS

Major Field: English

Biographical:

Education: Received Bachelor of Arts Degree in English from Khartoum University, Khartoum, Sudan in August 1975; received Master of Arts Degree in Modern English Language and Linguistics from Sheffield University, United Kingdom in May 1980. Completed the requirements for the Doctor of Philosophy degree with a major in Linguistics and TESL at Oklahoma State University in December 1995.

Experience: Taught English at high schools in Sudan; worked as an interpreter-translator for the Ministry of Defense and Aviation, Saudi Arabia for two years; employed by the University of Gezira, Sudan, as a lecturer in the English Language Center for two years; employed by the American Arabian Oil Company, Saudi Arabia as an Advanced Teacher of English at the Industrial Training Institute for two years; employed by the Foreign Language Department, King Faisal University, Saudi Arabia, as an instructor of English; employed by Oklahoma State University, Department of English as a graduate teaching associate, 1991 to May 1995.

Professional Membership: Oklahoma Teachers of English as a Second Language, Teachers of English to Speakers of Other Languages (TESOL), the Linguistic Society of America.