Original Article

INVESTIGATING LEXICAL COHESION IN DISCOURSE STUDIES: THEORETICAL INSIGHTS AND REAL-WORLD IMPLEMENTATIONS

Martin, P. T.

Department of Linguistics, University of Cape Town, Private Bag X3, Rondebosch, Cape Town, South Africa.

Abstract: This paper presents a comprehensive review of lexical cohesion studies in discourse analysis. The paper focuses on the theoretical, developmental and practical application of lexical cohesion in discourse studies. Cohesion is concerned with the linguistic resources on the surface of texts that signal relations between parts of texts. Its textual elements give texts their texture, making them powerful and effective pieces of language. The paper explores the emergence and nature of cohesion in discourse analysis, scholarly rhetoric, and debate on cohesion and textuality in discourse, approaches to lexical cohesion analysis, and insights from lexical cohesion studies. The review is structured into four major sections: (1) emergence and nature of cohesion in discourse analysis, (2) scholarly rhetoric and debate on cohesion and textuality in discourse, (3) approaches to lexical cohesion analysis, and (4) insights from lexical cohesion studies. The paper concludes by highlighting the significance of lexical cohesion as an approach to discourse analysis, revealing its potentials for analyzing different registers and genres (spoken and written) and how it facilitates the perception of relevance of the texts to the intended audience. Ultimately, the study seeks to contribute to the knowledge base of lexical cohesion by providing researchers and students an up-to-date overview of lexical cohesion as an approach to discourse analysis.

Keywords: Cohesion, Lexical Cohesion, Discourse Analysis, Coherence, Approach to Discourse Analysis, Approach to Lexical Cohesion, Lexical Cohesion Devices.

Introduction:

Halliday and Hasan's Cohesion in English (1976) brought lexical cohesion into the limelight and since then, cohesion has been a major source of inquiry among discourse analysts. Through the decades, lexical cohesion studies have made tremendous advances in terms of newer frameworks and exploring data, applying different models to investigate spoken and written registers and genres of discourse. Lexical cohesion studies have revealed how lexical resources interact with discourse goals, facilitate perception of textual relevance and contribute to the overall coherence of discourse. Despite criticisms of the cohesion theory, which argue that coherence is the only necessary property for the unity of texts, the significance of cohesive devices in texts cannot be overemphasized. This review synthesizes evidence from lexical cohesion studies to provide an up-to-date overview of lexical cohesion as an approach to discourse analysis. The four major sections of this review are emergence and nature of cohesion in discourse analysis, scholarly rhetoric and debate on cohesion and textuality in discourse, approaches to lexical cohesion analysis, and insights from lexical cohesion studies. Overall, this study illuminates the value of lexical cohesion as a tool for analyzing the texture and coherence of discourse, and how it can facilitate successful communication in authentic social interactions.

Original Article

1. Emergence and Nature of Cohesion in Discourse Analysis

Halliday and Hasan (1976) invoked the notion of cohesion in their attempt to account for the essential semantic relations between different elements on the surface of texts that enable the texts to 'hang together' as units of meanings. These text-forming elements, therefore, give texts their *texture* (Halliday and Hasan, 1985; Martin, 2001). Lack of these relations between the different elements in texts mostly renders them unintelligible, even after much struggle by the reader/listener, but cohesive texts are mostly found by readers/listeners as powerful and effective pieces of language (Eggins, 2004). On this basis, Halliday and Hasan (1976) argue that cohesion is a necessary and sufficient property for the unity of texts. To illustrate this point, they give the following examples:

(1) Wash and core *six cooking apples*. Put *them* into a fireproof dish. (Halliday and Hasan, 1976: 2) **(2)** Why does this little *boywriggle* all the time? *Girls* don't *wriggle*. (Halliday and Hasan, 1976: 285)

Example **1** above shows how *grammatical cohesion* is used to achieve connection between sentences. The pronoun *them* in the second sentence refers back (anaphorically) to the nominal group *six cooking apples* in the first sentence. The presupposition raised by the use of the pronoun *them* is therefore resolved by referring back to the presupposed referent *six cooking apples.* This type of relation is termed *Reference* by Halliday and Hasan (1976). On the other hand, example **2** depicts the use of *lexical cohesion* to establish the connection. In this case, two cohesive ties can be identified: *boy-girl* and *wriggle-wriggle* from the first and second sentences, respectively. The boy-girl relation is that of opposition, termed *Antonymy;* while the wriggle-wriggle relation is *Repetition,* which is simply a reiteration of the same lexical item. Therefore, both the sentences are cohesive by having these related resources.

1.1 Scholarly Rhetoric and Debate on Cohesion and Textuality in Discourse

It has been shown how Halliday and Hasan (1976) support their claim that cohesion is a necessary and sufficient property for the unity of texts. However, other linguists interested in discourse analysis have fiercely challenged Halliday's claim that cohesion is the necessary and sufficient property for the unity of texts. To these scholars, the only necessary property for the unity of texts is coherence – how the listener/reader perceives connection between the propositions expressed in the text (see, for example, Beaugrande and Dressler, 1981; Brown and Yule, 1983; Enkvist, 1978; Sanders and Maat, 2006; Widdowson, 1978; Yule, 1985). To begin with, Enkvist (1978), Brown and Yule (1983) and Yule (1993) all argue that a text may exhibit abundant cohesive signals and still fail to yield a unified whole. This is what Enkvist (1978) termed *pseudo-coherence* because even though the text is highly cohesive, the propositions expressed by the sentences are not mutually connected. These scholars gave examples of pseudo-coherent texts as follows:

(3) My father bought a *Lincoln convertible*. *The car* driven by the police was *red*. *That colour* doesn't suit her… (Yule, 1993:106)

(4) I bought a *Ford*. A *car* in which President Wilson rode down the Champs Elysees was *black*. *Black* English has been widely discussed… (Enkvist, 1978: 110)

The foremost rationale behind these scholars' argument is that the formal cohesive devices are not enough to guarantee *textness* because they alone cannot enable readers or listeners to make sense of what they read or hear. That we make sense of texts not because of the connections between the words and sentences but because of the perceptible coherence; therefore using cohesive elements in texts is a choice made by writers or speakers. However, these criticisms are obviously not enough to devalue the cohesion theory, because Halliday and Hasan (1976) have succeeded in drawing attention to some linguistic resources that contribute to the unity of texts. It is disillusioning to know that each of the examples given above was intuitively constructed to back the argument, when intuition is not enough to account for human textual interaction. At least, it must be seen that the longer an authentic text gets, the more likely is it to be cohesive (Tanskanen, 2006).

Nevertheless, some of these linguists also argue that texts can be coherent without being cohesive; that cohesion is an epiphenomenon of coherence, or an illusion evoked by the text coherence. In other words, they

agree that texts can achieve coherence even without the surface markers of cohesion. Widdowson (1978) and Widdowson (2004), for example, argue that discourse is not dependent on the overtly signaled cohesion. He insists that the most important property of texts is the *propositional development* and this is not always overtly signaled between sentences. He emphasizes that the propositions expressed in sentences are used to perform different illocutionary acts, and therefore listeners and readers can make sense of sentences by simply focusing attention on the illocutionary acts performed (by the propositions in the sentences) and not on the overt cohesion markers. To illustrate his points, Widdowson gives the following examples of non-cohesive but coherent utterances between interlocutors:

(5) **A:** What are the police doing?
B: I have just arrived. (Widdowson, 1978:27)

(6) **A:** That's the telephone.
B: I'm in the bath.
A: O.k. (Widdowson, 1978:29)

Similarly, Sanders and Maat (2006) insist that cohesion approach is inadequate because even without cohesive signals some texts would present no interpretation difficulties. They, therefore, believe that the overt textual connection is not necessary, and that connection in discourse is an attribute of mental representation of the text than the text itself. These scholars give the following example:

(7) The winter of 1963 was very cold. Many barn owls died.
(Sanders and Maat, 2006:592)

Each of the above excerpts (5, 6, & 7) shows no overt cohesion but interlocutors (or readers) can recognize the relationship between the propositions expressed or the illocutionary acts performed. This means the coherence of the texts is perceptible. Arguably, Widdowson's (1978) ideas are quite plausible, but it must be understood also that texts rarely achieve coherence without cohesion. Tanskenen, (2006) emphasizes that it is difficult to find real language data of certain length showing coherence without the surface cohesive resources. Consequently, these same created examples are mostly given to support this argument. For instance, Widdowson's constructed exchanges cited above have been used by scholars to support their argument (see, for example, Yule, 1993: 107; Lautamatti, 1990; Brown and Yule, 1983:196). It is therefore plausible to understand that texts may be coherent without cohesion, and that most coherent texts naturally exhibit cohesion.

Responding to these arguments, Hasan (1984) emphasized that coherence is a result of *cohesive harmony*. She observed that when cohesive chains interact, the result is cohesive harmony; and this consequently makes a text coherent. According to Hasan (1984), *chain interaction* is when at least two members of a chain are in cohesive relations with other members of different chains. She claimed that the denser the cohesive harmony in a text, the more coherent would be the text. Hasan (1984) added further that the phenomenon of coherence is to be judged only by the texts receivers. Similarly, some linguists concluded that for successful interpretation, both the text and the receiver's knowledge brought upon it must be taken into consideration. To illustrate this, Blackemore (1992) gives the following examples:

(8) The river had been dry for a long time. Everyone attended the funeral. (Blakemore, 1992:35)

(9) If a river has been dry for a long time, then a river spirit had died. Whenever a spirit dies there is a funeral. The river had been dry for a long time. Everyone attended the funeral. (Blakemore, 1992:36)

Blakemore gives these two examples to demonstrate how some texts would appear *irrelevant* to the audience for their lack of particular knowledge. From the perspective of Wilson and Sperber's (1986) Relevance Theory, Blakemore observes that a Westerner, who knows nothing about the beliefs of the *Sissala* people of Burkina Faso and Ghana, would have to expend great deal of cognitive effort in trying to process the assumptions in

text **8** (which visibly contains no cohesive elements). This is because he lacks the knowledge required to process the new assumptions in the text. On the other hand, text **9**, which visibly contains cohesive elements, is given to show how text 8 can be made *relevant* even to the Western audience who has known nothing about the Sissala tradition. In text **9**, the communicator has employed cohesive devices and this makes the input worth processing to the audience. From these examples, it can be understood, then, that use of cohesive devices in texts facilitates the perception of *relevance* of the texts to the intended audience. According to Wilson and Sperber (2006), the higher the cognitive effort required in processing an input, the lower would be its relevance, and vice versa. Therefore, text cohesion and its resultant perceived coherence that makes the text relevant have close relationship. From this, it can be easily conjured up that cohesion is an essential property in textual interaction.

Likewise, to Hoey (2005), cohesion correlates with coherence. He posits that the cohesive properties of words in texts are built into the words themselves. Reporting from the analysis of newspaper English, travel writing and literary texts corpus, Hoey establishes that while some lexical items are *primed* positively to participate in cohesion, others are negatively primed to avoid participating in cohesion. From his corpus, for example, he observed that *army, planet, year* and many other words are positively primed to participate in cohesive chains; while *asinine, wobble, blink* and some other words are negatively primed to avoid cohesion. Nevertheless, Hoey also suggests that the cohesively primed lexical items in texts are specifically primed to participate in certain kinds of cohesive relations and not others; and that they may be primed to participate in long chains, short chains, or only in links without chaining. From his data, for instance, Hoey discovered that while *planet* and *army* are primed to participate in chains of hyponymy*, ago* and *gay* are primed to participate in simple repetition. The linguist believes that *lexical priming* and cohesion ultimately influence the text organization. From these findings and points of view, it can therefore be seen that cohesion is not just an illusion or an epiphenomenon of coherence but an essential property in the creation of coherence (as Hasan 1984 also claims).

However, other linguists agree that any successful textual interaction requires the convergence of *cohesion, coherence,* and *the communicators* (see, for example, Gonzalez, 2011; Gonzalez, 2010; Tanskanen, 2006; Taboada, 2004). To these researchers, communicative situations demand cooperation and collaboration from both text producers and receivers. In either monologic or dialogic texts, explained these scholars, texts producers employ cohesive devices as signals of their collaboration towards coherence; and texts receivers also collaborate by assuming that those signals are provided by the texts producers for interpretive purposes. Therefore, these devices provided by the texts producers are expected to be collaboratively utilized also by the text receivers in interpreting the texts. This implies that the collaboration is realized at both the production and the interpretation stages of monologic discourses. At the production stage, texts producers mentally interact with the implied receiver and enact their discourses accordingly; and at the interpretation stage the receivers collaborate by looking for these collaborative signals built into the texts by the producers. Therefore, even if cohesion is not taken as a necessary and sufficient property for text unity, with these scholarly ideas and perspectives reviewed, one would certainly not hesitate to concede that it is a very essential property for achieving the coherence of texts. It contributes tremendously in facilitating the success of textual communication by easing the interpretation process.

2. Approaches to Lexical Cohesion Analysis

Withthe publication of the canonical Halliday and Hasan (1976) *Cohesion in English*, this particular research field was triggered not only on English language but also on other languages (see, for example, Enkvist, 1975 (Finnish & English); Danes, 1987 (Czech language), and different cohesion analysts have succeeded in coming up with different models of lexical cohesion. Lexical cohesion is the type of cohesive effect achieved by careful selection of lexical items that are related in different ways to other lexical items already used in the same discourse (Halliday and Hasan, 1976; Bloor and Bloor, 2004; Halliday, 1985/1994; Halliday and

Original Article

Matthiessen, 2014). However, there have been some controversies and conflicting views regarding the categories and relations identified by different lexical cohesion analysts. The aim here is to examine some of these models and highlight areas of their similarities, differences and controversies.

2.1 Halliday and Hasan's (1976) Model of Lexical Cohesion

To begin with, Halliday and Hasan's (1976) lexical cohesion model has two broad categories: *Reiteration* and *Collocation.* Different lexical relations have also been identified under each category. The lexical relations under reiteration include: Repetition, Synonymy, Near-synonymy, Superordinate, and General Words. Repetition is where a lexical item is repeated verbatim (exact repetition, like *ascent-ascent*), or with a slight morphological modification (in-exact repetition, like *boy- boy's*).Synonymy is a relation of lexical items with close (or similar) meanings, such as *ascent-clime*; near-synonymy is a relation between items whose meanings are nearly similar, such as *road-path*; superordinate is a relation between two or more lexical items in which the last item dominates the earlier one(s) in the lexical taxonomy, such as *ascent-task;* and general words are mostly common nouns and rarely indefinite pronouns that refer back (anaphorically) to presupposed items with which they have identical referents, such as *boy-(the)idiot, elm-(that)thing.* On one hand, the general words are mostly modified by reference items such as *the, this, and that,* which accompany them to signal the anaphoric function. They are, therefore, lexical but similar to the grammatical *reference.* On the other hand, the relations identified under the category of collocation include antonymy, hyphonymy, meronymy, co-hyphonymy, co-meronymy, ordered series, and co-occurrence tendency. Antonymy is a relation of opposition between items, such as *boy-girls;* Hyphonymy is where X is a type of Y, like *table-furniture;* Meronymy is a part-to-whole relation, like *car-brake;* Co-hyphonymy is a relation between members of the same general category, like *chair+table –furniture*; co-meronymy is a relation between parts of a common whole, like *wheel+brake – car;* Ordered Series is a relation between pairs of items belonging to the same set, like *Tuesday-Thursday, South-North, Colonel-Brigadier;* and Co-occurrence tendency accounts for words that regularly cooccur in adjacent contexts, like *joke-laugh, ill-doctor, try-succeed.* It is appropriate to note that in examining lexical cohesive relation, identity of reference between items is irrelevant. Moreover, in this framework, and of course in many others in our subsequent review here, grammatical items are ignored in lexical cohesion analysis, and also grammatical categories and morphological forms of lexical items do not restrict their relations with others.

However, the Halliday and Hasan's (1976) model of lexical cohesion has been criticized by different cohesion analysts. In particular, the use of the term *collocation* in the model is seriously challenged. The challengers insist that the term collocation is a lexicographic or lexicosemantic term by Firth (1957) meant for statistically examining the relation between *node* and its adjacent collocates (4 to 6 words away). They expatiate that cohesive relations in texts across clauses and sentences are relations of different kind, because items related collocationally in texts may not be so adjacent (Tanskanen, 2006; Lewin, Fine, & Young, 2001; Martin, 2001). The collocation in Halliday and Hasan (1976) model is also said to be vague and the lexical relations also loosely defined.

For this, Hoey (1991) describes it as a ragbag of lexical relations. Other lexical relations in this model are also said to be blurred. For example, Martin (1992) observes that Halliday does not provide any criteria for differentiating hyponymy from superordinate, and that general words are better considered grammatical than lexical because their roles in lexical cohesion is 'negligible' (Martin 1992:287). To Lewin et al. (2001), the relation of near-synonymy is difficult to define, and except for repetition, all the relations are loose because they largely depend on the subjective judgment of the text receiver. For these highlighted shortcomings in Halliday and Hasan (1976) model, most subsequent models avoided adopting some controversial terminologies and also re-organized the relations (although the material being investigated also influences the choice of terminologies and relations). But it is an indispensable fact that Halliday and Hasan (1976) have set the phase for this research field and most recent models draw upon this fundamental model. However, even

Original Article

Halliday and Hasan themselves have responded to some of these criticisms because each has revised the model in subsequent studies.

2.2 Hasan's (1984) Model of Lexical Cohesion

We turn now to Hasan's (1984) revised model of lexical cohesion, which has two broad categories: *General category* and *Instantial category.* Each of these two categories has some lexical relations identified under it. The lexical relations under general category include: repetition, synonymy, hyponymy, meronymy and antonymy. These relations are the same as those found in Halliday and Hasan (1976). On the other hand, the relations under instantial category, which are created by the texts under investigation, have the following relations: *equivalence, naming and semblance.* Equivalence is a relation between lexical items in particular texts where the items in question stand for the same referent, as in '*the sailor* was their *daddy*'; naming stands for relation between items where each names the same referent, as in '*The dog* was called *Toto*'; and semblance stands for a relation in which one entity is likened to another, as in '*The deck* was like *a pool'*. Hasan's model was influenced by the data being analysed (children's narratives). She is said to have chosen categories that suit her data best (Tanskanen, 2006). Therefore, this particular model is most effective in the analysis of narratives, but may not suit non-narratives (Hoey, 1991; Taboada, 2004; Tanskanen, 2006). Although this model also has two categories like Halliday and Hasan (1976), Hasan has avoided using the term collocation; and some relations previously handled under collocation (like meronymy, hyponymy, and antonymy) are now given fresh headings. Instead of collocation, Hasan uses *Instantial relations* which are not defined semantically but determined by specific contexts. It can also be seen that this revised model does not contain superordinate, near-synonymy and general words as obtained in Halliday and Hasan (1976).

2.3 Halliday's (1985) Model of Lexical Cohesion

Halliday's revised model (1985) has three general categories: *repetition, synonymy,* and *collocation*. For Halliday, repetition of a lexical item carries the strongest and most direct cohesive force that this particular cohesion should be given a unique status. Consequently, it is treated as an autonomous category in this model. In addition, identity of reference between cohesive elements is said to be irrelevant here too, so are morphological forms and grammatical classes. Therefore, *bear-(the) bear, dine-dining, strove-strife-strive* are considered cohesively related by repetition. Each pair is considered the same lexical item. The next category, synonymy, has many relations under it. The relations under synonymy in this model are further classed into two: those with identity of reference and those without identical referents. Synonymous lexical relations with identity of reference include: *Synonymy* (like blackbirds-birds) and *Superordinate* (like pig-creature), while those without identity of reference include: *hyponymy* (like chair + table – furniture), *Meronymy* (like bottle-stopper), *Co-hyponymy* (like plants + grass – vegetation), *Co-meronymy* (flowers-fountains= garden), and *Antonymy* (like asleep-woke). The last category in the model is *collocation*. The collocation relation here, like in Halliday and Hasan (1976), does not depend on any systematic semantic relation between items, but is based on *co-occurrence tendency.* An instance of this, according to Halliday, is the strong collocational bond between items like *smoke-pipe*, *snow-white*, and *cold-ice*.

The Halliday (1985) model is in many ways different from both Halliday and Hasan (1976) and Hasan (1984). However, it is closer to the former than the latter. In the first place, while Hasan (1984) drops the categorical label *collocation,* Halliday (1985) maintains this and also with similar meaning. In Hasan (1984), certain lexical relations used in Halliday and Hasan (1976) have been dropped, namely superordinate, general words, and near-synonymy; Halliday (1985) drops only general words and near-synonymy. All the three models identify the lexical relations: repetition, synoynymy, hyponymy, meronymy, and antonymy. Despite all the 'inadequacies' and 'vagueness' attributed to the use of collocation and superordinate (as reviewed above), Halliday (1985) still uses them again. Interestingly, most of the lexical relations identified in these models are generally similar.

Original Article

2.4 McCarthy's (1988) Model of Lexical Cohesion

It has been earlier highlighted that lexical cohesion models are often influenced by the materials being investigated. This was the case of Hasan (1984) and the children's narratives being analyzed. Similarly, the lexical cohesion model developed by McCarthy (1988) was also influenced by his natural conversation data. According to McCarthy (1988), the lexical cohesion model developed by Halliday and Hasan (1976) was not enough to handle the analysis of natural conversation data because it was meant for the analysis of written texts. McCarthy (1988) was interested in examining how interlocutors use lexical items in a systematic and patterned way for the interactive management of talk. Therefore, unlike earlier models, he focuses on cohesive relations within and across speaker turns, how speakers can relexicalize senses within their turns and those of others. Despite its renowned cohesive strength as shown in earlier models (including the Stratificational Gutwinski, 1976), *repetition* has not been included in McCarthy (1988) model. He argues that although repetition is very functional in conversations, it is more striking when non-identical forms are used to repeat contents; that this relexicalization makes conversational management among interlocutors most powerful. McCarthy's model also incorporates

Brazil's (1985a & b) *Communicative Theory of Intonation* to account for the role of intonation choices in signaling lexical relation. That to understand and to manipulate values of lexical items is discourse skill. This model identifies four categories of analysis: (a) change of item to retain same sense (equivalence), (b) change of item to increase sense (inclusion: specific-general), (c) change of item to include sense (inclusion: general-specific), and (d) change of item to oppose sense (opposition).

The McCarthy (1988) model is not close to Halliday and Hasan (1976) like Hasan (1984) is. While all Halliday and Hasan (1976), Hasan (1984) and Halliday (1985) employ repetition as a lexical relation, McCarthy (1988) does not. He claims that his model is more suitable for the analysis of conversation data than Halliday and Hasan (1976) is. McCarthy also opposes ~~to~~ the use of lexical relation labels like Synonymy, Antonymy, and so on because he believes they are not the best for capturing discourse-specific relations. For his choice of categories, McCarthy is said to be a phase setter among the promoters of the notion of discourse specificity in the analysis of texts (see, for example, Gonzalez, 2010, Gonzalez, 2011; and Tanskanen, 2006).

2.5 Hoey's (1991) Model of Lexical Cohesion

While McCarthy (1988) excluded repetition from his analysis, Hoey (1991) argues that lexical repetition has the optimum text-forming property. Therefore, in his lexical cohesion model developed for examining the patterns of lexis in non-narrative texts, the scholar included repetition. The importance attached by Hoey to repetition relations as uniquely essential properties of texts led to the inclusion of some grammatical elements. These grammatical elements (specifically pronouns), Hoey emphasizes, are included in the model because they allow text producers to 'say something again' as lexical repetitions do. Hoey reports that repetition relations in texts cue sentences that are most central to their meanings and therefore unravel their overall organizations. Despite his inclusion of grammatical elements in the model, Hoey observes that lexical cohesion is the most dominant in texts, and that the cohesive effect achieved by the grammatical elements is weaker. The categories in Hoey's model include: (a) *simple lexical repetition* (as in previous models, like *a bear-bears*), (b) *complex lexicalrepetition* (example *a drug- drugging*), (c) *simple paraphrase* (what in Halliday and Hasan, 1976 is named Synonymy, like *to sedate- to drug*), (d) *complex paraphrase* (what in Halliday and Hasan, 1976 is called Antonymy, like *heat-cold*), (e) *Substitution* (when a pronoun reiterates a noun, like *drug-it*), (f) *Co-reference* (when lexical items refer to an identical referent, as in Mrs. Thatcher – The Prime Minister), (g) ellipsis (when part of a lexical item is omitted in subsequent mention, as in *a work of art- the work*), and (h) *Deixis* (when a substitute item is used to refer to an entity already identified, as in *Plato and Aristotle - these writers*).

Hoey's (1991) model is unique in some ways. To begin with, for all the lexical cohesion models reviewed so far, Hoey's is the only one that includes grammatical elements. While the relation between an antecedent noun

and a pronoun referring back (anaphorically) to it is termed reference in Halliday and Hasan (1976), Hoey (1991) refers to this relation as *substitution.* Still, Halliday and Hasan (1976) refer to the relation between a pronoun (the presupposed) and another pronoun (the presupposing) as *reference,* but Hoey sees this as *repetition.* What most models refer to as synonymy is referred to as simple paraphrase in Hoey's model. Hoey's complex paraphrase is termed Antonymy by most models. What are termed simple and complex lexical repetitions in Hoey's model are both termed repetitions in Halliday and Hasan (1976). However, this model also focuses mostly on similar meaning relations as the previous models, except for terminological differences. Like Hasan (1984) and McCarthy (1988), this model also does not include collocation. The idea of *bonds* between sentences, which allow for identifying *central* and *marginal* sentences, contributes tremendously to our understanding of the text-forming properties of lexical items.

2.6 Martin's (1992) Model of Lexical Cohesion

It has been seen how Hoey's (1991) model excludes collocation. Other models, however, simply attempt to precisely re-define the semantic relations earlier treated under collocation. A good example of these is Martin (1992). The Martin's lexical cohesion model was constructed to target the analysis of texts of given fields (like tennis, solar system and education). He argues that lexical items in texts are organized differently in accordance with the given fields. The lexical relations identified in this model are grouped into three major categories: *taxonomic*, *nuclear*, and *activity sequence*. Interestingly, the last two categories here (*nuclear* and *activity sequence*) are simply modifications of collocation as obtained in Halliday and Hasan (1976). To examine each of these categories carefully, the relations identified under taxonomic are further grouped into *superordination* and *composition*. The relations under superordination taxonomy are based on sub-classification (the 'is a' relation), while those under composition taxonomy are based on part-to-whole (the 'has a' relation). In sum, the relations under taxonomic include: hyponymy, co-hyponymy, contrast, synonymy, repetition, meronymy, and co-meronymy. Martin observes that people, places, and things would mostly be organized in accordance with the types of taxonomy in those fields. On the other hand, nuclear relations are based on how actions, people, places, things and qualities would configure as activities in activity sequences from different fields (from tennis, examples include *volley-winner*, *lobreturn, smash-overhead* and so on). While activity sequences are based on expectancy relations between the items in activity sequences from given fields (like *opponent-lobs, player-smashes, opponent-retrieves,* and so forth). Cohesion analysts employing this model are expected to conduct their analyses in three phases, to be able to capture all the taxonomic, nuclear, and activity sequence relations in their data (Martin, 1992).

Martin's (1992) model is in some ways similar to some models reviewed above. It is also different in some ways from other models discussed so far. For instance, there are striking similarities between Martin's taxonomic relations and Hasan's (1984) general relations, and also Halliday's (1985) synonymy relations. However, Martin's model is different from Halliday & Hasan (1976) and Hasan (1984) in that collocation is not used. Instead, the lexical relations previously handled under the collocation category have been unpacked and given precise definitions under nuclear and activity sequences. Unfortunately, analyses using this model become too cumbersome because three distinct steps must be taken, one for each of the categories. For this, the model is not suitable for the analysis of longer texts (Tanskanen, 2006), and cohesion analysts are sometimes concerned with long texts.

2.7 Taboada's (2004) Model of Lexical Cohesion

While Martin (1992) was concerned with written texts, Taboada (2004), like McCarthy (1988), was also concerned with the analysis of natural conversation data. But unlike McCarthy (1988), who thought that Halliday and Hasan (1976) model was inadequate in handling conversational data, Taboada's (2004) model was largely an adaptation of the Halliday and Hasan (1976) model. Taboada's focus was to compare English and Spanish conversations with the aim of unraveling the resources speakers utilize in building coherent conversations. She was, therefore, interested in the text-forming discourse characteristics that enable the

Original Article

dialogues hold together. Taboada's study employed three frameworks of analysis: *Speech Genre* (Bakhtin, 1986), *Rhetorical Structure Theory* (Mann & Thompson, 1988), and *Cohesion theory* (Halliday and Hasan, 1976), but our concern would be limited here only to the cohesion theory. Taboada (2004) also examined both grammatical and lexical cohesion but the relations are slightly modified. While Halliday and Hasan (1976) identified four grammatical cohesive relations (repetition, substitution, ellipsis, and conjunction), Taboada excluded *conjunction* from her analysis. Coming to lexical cohesion, Taboada (2004), like Halliday and Hasan (1976), also identified two broad categories: *Reiteration* and *collocation*. The relations identified under reiteration include: *same item* (exact & rephrased), *synonymy, superordinate, subordinate, and general words.* Collocation, as in Halliday and Hasan (1976), is based on association of items that usually co-occur.

However, it is obvious that the Taboada's (2004) model of lexical cohesion is also slightly reformulated for the purpose of her analysis. In the first place, what is called *repetition* in Halliday and Hasan (1976) is termed *same item* in Taboada, and more importantly, it has been broken into two: *exact* and *rephrased.* The exact repetition here is the same as *repetition* obtained in Halliday and Hasan (1976), but rephrased is a subcategorization established by Taboada. *Subordination* is also Taboada's invention, Halliday and Hasan (1976) do~~es~~ not include this. But synonymy, superordinate, general word and collocation are the same as that obtained in Halliday and Hasan (1976). So while most cohesion analysts (except for Halliday, 1985/1994; Halliday and Matthiessen, 2014) avoid using the controversial *collocation*, Taboada employs it in her model; and while some cohesion analysts think Halliday and Hasan (1976) miss much to be capable of dealing with conversation data, Taboada (2004) believes it can be employed and slightly modified to function efficiently as such.

2.8 Tanskanen's (2006) Model of Lexical Cohesion

Like Taboada (2004), Tanskanen (2006) model of lexical cohesion was also an adaptation of Halliday and Hasan (1976). But while Taboada was concerned with the analysis of natural conversation data, Tanskanen was concerned with the analysis of different text types (spoken and written), and this has obviously influenced the choice of relations in the model. Tanskanen posits that cohesion is a resource that communicators utilize in (collaborating towards coherence) achieving coherence. She was interested in examining the variations in the use of cohesive devices that result from different conditions of texts production. In addition to Halliday and Hasan (1976) model, Tanskanen's model has also been influenced by many other models like McCarthy (1988), Hoey (1991) and Martin (1992). As in Halliday and Hasan (1976), this model has two broad categories of lexical relations: *Reiteration* and *Collocation.* Eight different lexical relations are identified under reiteration, and they are: *simple repetition* (exact items or with simple grammar change), *Complex repetition* (based on different grammatical function), *substitution* (where an item substitutes another), *equivalence* (what is termed synonymy in some models), *Generalisation* (specific-to-general relation, like superordinate in some models), *Specification* (general-to-specific relation, like meronymy), *co-specification* (like co-meronymy or co-hyponymy), *contrast* (Anonymy in some models). On the other hand, three lexical relations are identified under collocation: *Ordered set* (like days of the week or months of the year), *activity-related collocation* (related in terms of activity, like *car-drive, meal-eat*) and *elaborative collocation* (co-occurrence tendency).

Many features of Tanskanen's (2006) model are worth commenting on. Firstly, while many cohesion analysts have avoided the inclusion of the notorious collocation in their models, Tanskanen included it and redefined it by drawing on Martin's (1992) *nuclear relations.* From Hoey (1991), Tanskanen borrowed the idea of including some grammatical items in her model, and also that the relation between a noun and pronoun is *substitution,* and the repetition relation between pronouns. From McCarthy (1988), Tanskanen's model got the idea that lexical relations are created and controlled by the particular texts, that each text may make its unique meaning; therefore, analyses should be discourse-specific. This is why the terms used in the model are not from lexical semantics. This model is interesting but too complex because the relations identified are very many. It would visibly not be convenient for handling long texts but from the perspective of discourse, the

model is so versatile because it allows for the analysis of cohesion in different texts types (Flowerdew, 2013). Tanskanen herself has acknowledged the relatively high number of relations in the model but wishes that this would not make it too complicated than necessary. However, researchers conducting similar studies would find the model quite interesting and adaptable.

2.9 Gonzalez's (2010) Model of Lexical Cohesion

As in the case of Tanskanen (2006) and McCarthy (1988) models, the idea of discourse-specificity in analyzing lexical relations in texts also underlies Gonzalez's (2010) & (2011) model of lexical cohesion. To Gonzalez too, meaning relations between lexical items in texts are context-specific, that lexical cohesion analysis should focus on the communicative potentials of lexical items not their lexico-semantic meaning potential. Therefore, in her studies on lexical cohesion in telephone conversations (2010) and multiparty conversations (2011), the researcher offers an integrative model of lexical cohesion analysis. In both these studies, the texts were the starting points of the analyses and more emphasis was given to the context-specific than the decontextualized meanings of lexical items. The model has five meaning relations: *Repetition, synonymy, Opposition, Inclusion, and Associative cohesion.* Obviously, the new thing about this model is the *associative cohesion.* This is a replacement of Halliday and Hasan's (1976) collocation, so that the model would be capable of accommodating the analysis of instantial relations between items in specific texts without necessarily being *collocates* decontextually. It concerns with associative relations that operate across stretches of texts, both within and across speaker turns. Therefore, associates in texts need not be collocates.

Although the Gonzalez (2010) model, which is also employed in Gonzalez (2011), was also constructed for the analysis of conversation data like McCarthy (1988), it can be seen that the former has included repetition among the relations while the latter has not. While McCarthy has emphasized intonation choices made by speakers in determining cohesion, Gonzalez has not. It has been earlier highlighted that McCarthy (1988) contends that Halliday and Hasan (1976) model misses much to be capable of the analysis of conversation data, because it was meant for written texts. But it can be seen that the choice of lexical relations in Gonzalez (2010) is closer to Halliday and Hasan (1976) than McCarthy (1988). However, certain things are clearly similar between Gonzalez (2010) and McCarthy (1988) models: both are meant for conversation data analysis, both are based on discourse specificity of lexical relations, and both have relatively fewer relations compared to most models of lexical cohesion analysis (reviewed so far).

2.10 Halliday and Matthiessen's (2014) Model of Lexical Cohesion

While lexical cohesion analysts are busy innovating terms and reformulating earlier models to suit the nature of their data (see, for example, Gonzalez, 2011; Gonzalez, 2010; Tanskanen, 2006; Taboada, 2004; Hoey, 1991; McCarthy, 1988; Hasan, 1984 reviewed here), Halliday and Matthiessen (2014) offer a radical revision of Halliday, (1985/ 1994) model. This revised model has three broad categories of relations: *Elaborating relations, Extending relations, and Collocation.* It could be recalled that Halliday (1985) model also had three categories: *Repetition, Synonymy* and *Collocation*; where most of the relations were identified under Synonymy, while repetition and collocation stand alone as autonomous relations. Conversely, in Halliday and Matthiessen (2014), the category of elaborating relations comprises of *repetition, synonymy* (with identical referent 'synonymy proper' 'superordinate' as in *ankylosaur-creature* or without identical referent 'Antonymy' as in *woke-asleep*), and *hyponymy* (involving comeronymy). Extending relations involve only *meronymy* (involving co-hyponymy); and *collocation* takes care of lexical relations not depending on any systematic semantic meanings but the co-occurrence tendency of the items (as in *dine-restaurant, fry-pan*).

Therefore, regarding the lexical relations identified in each model, there is little difference between Halliday (1985) and Halliday and Matthiessen (2014), and even the controversial collocation is maintained. That means while some researchers feel that collocation is not a suitable term for cohesion analysis, Halliday and Matthiessen (2014) still see it as the best term to employ, unlike Hasan (1984) where collocation was avoided. Halliday and Matthiessen (2014) model is still closer to Halliday and Hasan (1976) than Hasan (1984), because

the same lexical relations have also been maintained; only that in Halliday and Matthiessen (2014) many lexical relations have been assigned to different categories than in Halliday and Hasan (1976) model. For example, in Hasan (1984), meronymy, co-meronymy, hyponymy and antonymy are all under synonymy, but in Halliday and Mattiessen (2014) meronymy and co-meronymy are under extending relations; while hyponymy and antonymy are also under elaborating relations. The same relations but classified differently.

2.11 Eggins' (2004) Model of Lexical Cohesion

This model comprises two broad categories: *Taxonomic lexical relations* and *Expectancy relations.* The relations under taxonomic include: *co-hyponymy, class/sub-class, contrast, synonymy, repetition, meronymy and co*-meronymy; while *expectancy relations* are based on the co-occurrence tendency of lexical items. The lexical relations identified in this model are discussed as follows:

1) *Co-hyponymy*–this relation is the same as found in Martin, (1992), Halliday (1985/1994), and Halliday and Matthiessen (2014). It is the relation between two or more lexical items where each is a *subordinate* member of a common *superordinate* class. An example is how *mango*, *banana*, and *orange* relate as co-hyponyms of the superordinate term *fruit*. Eggins' (2004) framework identifies this as a cohesive relation, and co-hyponymy relation between lexical items remains constant even when the superordinate item is not found in the text.

2) *Class/sub-class* –this particular relation is termed *hyponymy* in some lexical cohesion frameworks, like Martin, (1992), Halliday, (1985/1994), Halliday and Matthiessen, (2014), and Hasan, (1984). McCarthy (1988) terms it*inclusion* (general specific), Tanskanen (2006) calls it *generalization,* and Taboada (2004) terms it *superordinate.* Eggins's (2004) model uses the term to refer to a lexical relation in which items are related through sub-classification, where an item standing for a general class is related to another for being a sub-class member of the general class; the X is a type of Y relation. Examples include: car/jaguar, dog/greyhound, and furniture/chair. However, this relation may hold between one general item and one or more sub-class items.

3) *Contrast* –this is the relationship of opposition between lexical items. It has been termed differently in different models: as *Antonymy* (like Halliday and Hasan, 1976; Hasan, 1984; and Halliday and Matthiessen, 2014), as *opposition* (like McCarthy, 1988; Gonzalez, 2010; and Gonzalez, 2011), and as *complex paraphrase* (Hoey, 1991). Eggins (2004), like Martin (1992) and Tanskanen (2006), refers to this relation as *contrast.* Worthyof particular note here is the fact that items relating by contrast need not be considered antonymous in the lexical semantic sense. Contrastiveness of lexical items is determined by the context or text. This is not to mean that the items may not also relate contrastively from the semantic or decontextualized perspectives, but this is not a necessary condition. Examples include: strong/weak, out of fashion/up to date, old-aged pensioners/ the working people.

4) *Synonymy* –this is the relationship of similarity of meaning between lexical items. The relation is found in almost all models of lexical cohesion, as far as the researcher knows. It is termed *synonymy* in most of these models (see, for example, Halliday and Hasan, 1976; Hasan, 1984; Martin, 1992; and Taboada, 2004), but McCarthy (1988) and Tanskaken (2006) term it *equivalence,* and Hoey (1991) labels it *simple paraphrase.* The Eggins' (2004) model also terms it synonymy. Examples include: *happy/glad*, *cavalry/horses*, and *the Nazi extermination of the Jews/the Nazi slaughter.*

5) *Repetition* –as the name implies, this is a relation between items in which the same form, irrespective of identity of reference, is subsequently mentioned. Interestingly, the morphological form of the subsequent item might be slightly modified, or it may even belong to a different category than the earlier-mentioned item. Some frameworks differentiate further between *exact* and *inexact repetition* (see, for example, Halliday and Hasan, 1976; Taboada, 2004; Gonzalez, 2010; and Gonzalez, 2011) while others have *simple* and *complex repetition* (like Hoey, 1991; and Tanskanen, 2006). However, McCarthy's (1988) framework excludes repetition as a lexical relation. The model focuses more on how interlocutors feed in new items into the

conversations. Eggins (2004) includes repetition in her framework, and here too, identity of reference between the items related is not a necessary criterion. The relation is found in most frameworks of lexical cohesion (see, for example, Gutwinski, 1976; Halliday and Hasan, 1976; Hasan, 1984; Hoey, 1991).

6) *Meronymy* –this is the part-to-whole (or vice versa) relation between lexical items in texts. Examples include: tree/branch, bottle/stopper. Most frameworks have this relation, and it is similarly termed *meronymy* in most frameworks (see, for example, Halliday, 1985/1994; Hasan, 1984; Martin, 1992; Halliday and Matthiessen, 2014) but others term it differently, such as McCarthy (1988)and Gonzalez (2010), and Gonzalez (2011) term it *inclusion: general-specific*; while Tanskanen (2006) terms it *specification.* Other researchers, such as Hoey (1991) and Taboada (2004), exclude it from their frameworks of analysis. The meronymy relation is also included in Eggins' (2004) framework because it contributes greatly in the cohesion of texts.

7) *Co-meronymy* –the relation between items by being parts of a common whole, like the *hand/eye –body*, *branches/leaves –tree, flowers/fountains –garden* relation. As in the case of co-hyponyms, it is not necessary to find the general items in the texts displaying the co-meronymy relation. For example, *hand* and *eye* are co-meronyms even when the text in which they are used does not contain *body.* Most frameworks have this relation (see, for example, Martin, 1992; Halliday, 1985/1994; Halliday and Matthiessen 2014) but it is termed differently in some models;~~like~~ Tanskanen (2006) terms it *co-specification.* Only a few models, like Hoey (1991) and Taboada (2004) do not have it.

8) *Expectancy relation* –this category of lexical relations is not far from Halliday and Hasan's (1976) *collocation* or Gutwinski's (1976) *co-occurrence group* that are based on relations between lexical items that often co-occur in contexts. Examples include *train/track, baggage/car* and so on. Because the term *collocation* in some frameworks (like Halliday and Hasan, 1976; Halliday, 1985/1994; Taboada, 2004; and Halliday and Matthiessen, 2014) has been challenged for its 'subtlety' and 'vagueness' in the analysis of texts, many cohesion analysts have avoided including it in their frameworks without being redefined (see, for instance, Martin, 1992; Tanskanen, 2006; Gonzalez, 2010 and Gonzalez, 2011). However, some frameworks completely exclude collocation (see, also, Hasan, 1984; McCarthy, 1988; and Hoey1991). Therefore, Eggins' (2004) expectancy relation is a redefinition of Halliday and Hasan (1976) collocation. In this framework, expectancy relations hold between lexical items that *go together* based on any of the following four major conditions:

a. between an action and the characteristic (or expected) doer of the action; examples include: *doctor/diagnose,*
research/discover, police/arrest

b. between an action/process and the characteristic sufferer affected by the action; examples include: *play/guitar, read/book, cook/soup*

c. between an event/process and its typical location of occurrence; example include: *learn/school, read/library, cook/kitchen*

d. between compound nouns and the individual lexical items forming their parts; examples include: *heart/disease,*
child/birth

Finally, from our review of models of lexical cohesion above, it is quite logical to conclude that the similarities between these models outweigh their differences. For this, unless for terminological differences, all these models are concerned with similar lexical relations such as hyponymy, meronymy, antonymy, and so forth. However, major differences involve the use of the term collocation, where only a few models employ it; and more importantly the discourse-oriented approach to lexical relations employed in some models (see McCarthy 1988; Tanskanen, 2006; Gonzalez 2010; and Gonzalez 2011). In any case, it is plausible to understand that cohesion analysts develop models that are most suitable for the data being handled. In these

attempts, these researchers inevitably employ modified models (see, for example, Hasan, 1984; McCarthy 1988; Taboada, 2004; Tanskanen, 2006; Gonzalez, 2010; and Gonzalez 2011).

3. Insights from Lexical Cohesion Studies

This section focuses on findings from different lexical cohesion studies. Many text analysts have employed the cohesion theory to explore how this text-forming property operates in written and spoken texts, or how text producers utilize cohesive resources in different ways to achieve coherence and other characteristics of discourse (monologically or dialogically). Different findings have emerged from different cohesion studies.

3.1 Insights from Lexical Cohesion Studies on Written Discourse

A study by Morris and Hirst (1991), which focused on lexical cohesion in texts done by computational linguists, sought to experiment the analysis of lexical cohesion using computer programs. The study adapted Halliday and Hasan's (1976) lexical cohesion model. More importantly, the researchers also focused on examining the predictability relations between lexical cohesive chains, coherence, and discourse structure. However, the study revealed that not all lexical relations in texts can be detected by computer programs, especially the non-systematic semantic relations such as collocation, activity-sequence, and expectancy, and the discourse-specific relations; relations based on systematic semantics can be well detected. They also reported that lexical chains correspond to topic continuity in texts. Finally, similar to the findings of Morris and Hirst (1991), McCarthy (1988) and Tanskanen (2006), argued that text-based knowledge is obligatory in any attempt to account for non-systematic lexical relations in texts. Therefore, lexical cohesion analyses, especially those adopting the idea of discourse specificity of lexical relations, cannot be handled by computer programs alone. The analysis must involve both world-knowledge and text-knowledge. Morris and Hirst's findings that lexical cohesive chains correspond to topic continuity in texts imply the fact that lexical cohesion contributes to the overall coherence of texts.

While Morris and Hirst (1991) were concerned with the relation between lexical cohesion, coherence, and the structure of texts, Lewin, Fine, and Young (2001) were concerned with the relation between lexical cohesion and moves in the different genre units of Social Science Research (SSR) articles. Specifically, the researchers focused on the Introduction and Discussion sections as genre units in the SSR research articles. The study discovered no correspondence between lexical cohesion and moves, but that researchers preponderantly employ *repetition and synonymy* (99%) in both the introduction and discussion sections of the articles. This, therefore, points to the facts that scientific genres tend to utilize resources meant for clarity and definition, and more importantly, that texts of the same genre seem to exhibit the same cohesive features. Accordingly, studies on narrative texts such as Hasan (1984), where precision would not be so significant, for example, have come up with different patterns of lexical cohesion (Lewin et al 2001). Therefore, from the findings of this study, lexical cohesion is understood to be much related to genre, but may not be so related to the moves in the genre. Exploring lexical cohesion in different genres, text analysts also examine cohesion in media discourse. Hameed (2008) examined cohesion in magazine news articles. The study adapts Halliday and Hasan's model (1976), and was concerned with both grammatical and lexical cohesion. The researcher sought to identify the cohesive features of the texts, and to unravel the dominant cohesion. On his findings, Hameed reported that lexical cohesion, in the form of repetition, synonymy, and collocation, was the dominant cohesion in the magazine news articles. Like Morris and Hirst (1991) and Tanskanen (2006), Hameed's study also discovered that lexical cohesive chains corresponded to the topic development in the texts. This preponderance of repetition and synonymy as cohesive resources in magazine news article corresponds to Lewin et al (2001) findings on the lexical cohesion in the SSR research articles. Therefore, as also argued by Hoey (1991), lexical cohesion is the most dominant cohesion in these text types.

While discussing lexical cohesion in written texts, let us also take the study of Mirzapour and Ahmed (2011), which is a comparative study of lexical cohesion in English and Persian research articles. The study adapted

Original Article

Halliday and Hasan's (1976) model of lexical cohesion and it focused on examining the degrees of utilization of the different types of lexical cohesion in the articles. The data comprised 60 articles (30 English, 30 Persian) drawn from linguistic, Literature, Library and Information disciplines. The findings of the study indicated that the most frequent lexical cohesion in English research articles included repetition, collocation, synonymy, general nouns, meronymy, and antonymy; while those in Persian included repetition, synonymy, collocation, antonymy, hyponymy, meronymy, and general nouns. In both English and Persian research articles, the most frequent lexical relations used were repetition, collocation, and synonymy. While collocation was more frequent than synonymy in English, reverse was the case in Persian. However, the findings of this study correspond to that of Lewin et al (2001) and Hameed (2008) where repetition, synonymy, and collocation were found as the most preponderant cohesion. Many written texts analyses also reported that repetition is the most frequent cohesion (see, for example, Hoey, 1991; Hameed, 2008; Lewin et al, 2001; Liu & Braine, 2005; Mohammed-sayidina, 2010).

Another cohesion study is by Jabeen, Faiz, Mehmood and Yousaf (2013), who explored the relation between the cohesive devices and semantic qualities of written stories. It adapted Halliday and Hasan (1976) model, and both grammatical and lexical cohesion were examined. On lexical cohesion in the written stories, the researchers found that synonymy, near-synonymy, repetition, and collocation were mostly employed to create relation between different elements in the texts, and this essential feature enabled the whole text to be unified as a meaningful unit.

Therefore, like many other cohesion analyses, this study has shown how cohesive devices in texts enhance the perception of coherence in the texts. Although it dealt with a different genre, the study revealed similar patterns of lexical cohesion to those identified by earlier written texts analyses (see, for example, Hameed, 2008; Lewin et al, 2001; Mirzapour& Ahmad, 2011).

Finally, Malah (2015) explored lexical cohesion in Applied Linguistics research abstracts. The focus of his study was to unravel the patterns of lexis typical of these abstracts, and also to examine the extent to which the lexical ties contribute to the generic coherence of the abstracts. The study drew on Halliday and Hasan's (1976) model of lexical cohesion, and the research articles sampled were drawn from Discourse Analysis, Second Language Acquisition, Critical Discourse Analysis, and Contrastive Linguistics. The data were 40 research articles abstracts of 7,660 words. The analysis revealed 754 lexical ties, where the most frequent were *Repetition* (54%), *Collocation* (14%), and *Hyponymy* (11%). The study also discovered that lexical cohesion contributed to the attainment of the overall generic coherence of the abstracts. Therefore, this study reported similar findings to those of Hoey, (1991), Mirzapour and Ahmed (2011), Lewin et al. (2001), Jabeen et al. (2013), which also reported that repetition, collocation, and hyponymy are the most dominant in their data, and that lexical cohesion contributes to the coherence of the texts.

3.2 Insights from Lexical Cohesion Studies on Spoken Discourse

The relation between lexical cohesion and genre does not exist only in written genres. Angermeyer (2002) explored the phenomenon of lexical cohesion in multilingual conversations. The researcher was interested in examining how multilingual interlocutors utilize lexical cohesion in their conversations. The data for this study was drawn from recorded conversations of trilingual family of English, French and German in Canada. The study reported that lexical insertions, which could be even from previous conversation episodes, from the other languages into the matrix language were related to the conversational structure. More importantly, these lexical insertions enable the multilingual speakers to achieve cohesion in their utterances. The researcher claims that lexical insertions (in the corpus) are results of the multilingual speakers' attempts to create coherence between utterances in different languages. Drawing on Halliday and Hasan (1976) and Hoey (1991) models, he further argues that intrasentential code-switching can be explained as a result of the multilingual speakers' intention to make their utterances contextually coherent. Based on the findings of Angermeyer (2002), lexical cohesion in multilingual conversations also takes the form of lexical insertions, where speakers

in this hybrid genre utilize lexical resources from different languages to achieve cohesion. It is also shown that insertions, like intersentential code-switching, contribute greatly to the overall structure of multilingual conversations. The study could be said to reveal that lexical cohesion contributes to the coherence and also the overall structure of multilingual conversations, such as Morris and Hirst (1991) found in written texts.

Similar to Angermeyer's (2002) study, Taboada's (2004) also examined cohesion in conversations. But while Angermeyer (2002) was concerned with multilingual conversations of English, French and German, Taboada (2004) was concerned with English and Spanish conversations. She developed interest in diagnosing and comparing the resources that speakers utilize in *building coherence and cohesion*. Taboada's (2004) study adapted a modified model of Halliday and Hasan (1976), and was concerned with both grammatical and lexical cohesion. The researcher observed that lexical cohesion was the most dominant cohesion in both English and Spanish conversations. Exact repetition was the most dominant cohesion in both the languages, which was followed by inexact repetition and superordinate in English; and collocation and inexact repetition in Spanish, respectively. But Spanish used a higher number of links (620), and of course higher lexical links, than English (464), while both used exactly the same ratio of ties per word. In the corpus for this study, it was discovered that the cohesive chains hardly interacted, but the conversations were well coherent. Therefore, this study has presented findings that are quite in agreement with some earlier cohesion analyses, and also in disagreement with others. For instance, the findings of this study agree with Hoey (1991) where lexical cohesion was discovered to be the most dominant source of cohesion and also the preponderance of repetition. On repetition as the most dominant lexical cohesion, similar finding was reported by Lewin et al. (2001) study on lexical cohesion on SSR research articles. However, the findings of this study that cohesive chains hardly interacted in her coherent texts, disagrees with Hasan (1984), which claimed that coherence is a result of cohesive harmony, brought about by chain interaction. In other words, Hasan (1984) observed that coherence in texts is determined by the interaction of cohesive chains in the text that leads to cohesive harmony. Commenting on this conflict, Taboada suggested that different measures of cohesive harmony are required for different text types.

While some cohesion analyses like Taboada (2004) attempt to compare texts of the same genre, others compare texts of different genres. Tanskanen's (2006) comparison of texts of different genres fetches the study unique credit among discourse analysts (Flowerdew, 2013).

Tanskanen considered lexical cohesion as a resource utilized by communicators to achieve coherence. Consequently, the study employed a revised model of Halliday and Hasan (1976) to explore lexical cohesion in different text types: *face-to-face conversations, prepared speeches, electronic mailing lists, and academic writings.* However, interesting findings emerged from this study: that in all text types, reiteration and collocation participated in the cohesive chains (short and long), but collocation was comparatively lower than reiteration; that in all the texts investigated, the span of cohesive chains corresponded to those of topic segments; that two-party conversations were the most cohesive types of texts (160 ties) while academic writing had the least cohesive types of texts (105 ties). Based on these findings, the researcher concluded that collocation relation is relatively rare in texts, and that two-party conversations are more cohesive than academic writings. This study has similar findings with the study of Morris and Hirst (1991) on the correspondence between spans of cohesive chains and topic segments in texts. It is also in line with the findings of Hasan (1984), Hoey (1991), and Morris and Hirst (1991) that lexical cohesive chains contribute to the coherence of texts; and Angermeyer (2002) and Taboada (2004) that lexical cohesion operates in conversation. Most importantly, the study has revealed that lexical cohesion is a phenomenon that cuts across texts of all genres (or at least the genres examined).

Unlike the case of Tanskanen (2006) who was concerned with different text types, Wu (2010) was concerned with spoken monologic texts only. Wu's study investigated the relationship between lexical cohesion and the quality of oral English produced by Chinese undergraduates. The study was based on Halliday and Hasan

Original Article

(1976) model of lexical cohesion. Wu reported that there was high relation between lexical cohesion and oral English quality. The study reported that High Quality Discourses (HQDs) tend to utilize more lexical cohesion than Low Quality Discourses(LQDs). The two differed qualitatively and quantitatively in their use of lexical cohesion; the more the quantity of lexical cohesion employed, the higher the quality of the students' oral English. The findings also demonstrate that repetition and general nouns were the most frequent cohesive devices employed in both HQDs and LQDs, which are in agreement with those of other studies. The fact that lexical cohesion contributes to the quality of texts is not far from the fact that it remains as the dominant source of cohesion in most texts as also reported by Hoey (1991), Hameed (2008) and Tanskanen (2006). Other cohesion studies, such as Taboada (2004), Gonzalez (2010), and Gonzalez (2011) reported that repetition is the most dominant cohesion in spoken genres.

Another lexical cohesion analysis on spoken genre is the study of Gonzalez (2010) on telephone conversations, which examined how lexical cohesion operated in telephone conversations, and how interlocutors utilize lexical cohesive devices in achieving coherence and other generic characteristics of telephone conversations. The study employed an integrative model based on the model of Halliday and Hasan (1976), and the data was drawn from the *International Corpus of English Great Britain.* In total, 15 telephone conversations of 20,043 words were analyzed. In this study too, ties were established both within and across speaker turns. Like McCarthy (1988), Tanskanen (2006), Morris and Hirst (1991), Gonzalez also adopted the discourse-specific approach in this study. Therefore, in her analysis, lexical relations are determined by the system of the text and not the system of the language. More interestingly, Gonzalez succeeded in coming up with rich findings from this study. She discovered that, like face-to-face conversations (see, for example, Angermeyer, 2002; Taboada, 2004; & Tanskanen, 2006) or written texts (see, for example, Hoey, 1991; Hameed, 2008; Mirzapour& Ahmed, 2011; Hasan, 1984; Morris and Hirst, 1991), telephone conversations are also lexically cohesive, which contributes to their coherence and generic features. Gonzalez observed that the dominant kinds of lexical cohesion in telephone conversations are repetition, associative cohesion, and inclusive relations. She claimed that repetition, synonymyand opposition are the mostly utilized by interlocutors as topic continuity devices, while associative cohesion and inclusive relations are largely employed to shift focus or drift to different aspects of global topics. She also reported that remote-mediated ties enable interlocutors to collaborate different segments of conversations either within or across turns (WT and AT). Therefore, Gonzalez's (2010) integrative model of lexical cohesion, similar to those of Hoey (1991), Martin (1992), Hasan (1984), McCarthy (1988), Tanskanen (2006) and Eggins (2004), was successfully employed in this analysis. Similarly, other studies such as Morris and Hirst (1991) Tanskanen (2006) Hameed (2008), Gonzalez (2010) discovered that lexical cohesive chains correspond to topic segments in texts.

From the analysis of telephone conversations in 2010, Gonzalez shifted her attention to broadcast multiparty conversations in 2011, which examined the interaction between lexical cohesion, coherence, and the genre characteristics of multiparty conversations. The same integrative model of lexical cohesion as used in Gonzalez (2010) analysis of telephone conversations was employed in this study. Like Gonzalez (2010), McCarthy (1988) and Tanskanen (2006), this study also adopted discourse specificity in analyzing lexical relations, where decontextualized relations between lexical items in texts are not given much attention than the contextualized relations.

Her data were seven broadcast discussions (5 radio and 2 TV) of 15,683 words extracted from the public conversation category of the *International Corpus of English Great Britain.* The findings of her study indicate that because broadcast discussions were opinionative conversations constrained by the production limitations of live performances, which are anchored by chairpersons and intended by for an audience, they are highly lexically cohesive. The most dominantlexical cohesions discovered include repetition (59%), associative cohesion (24%), and inclusive relations (8.2%). By their nature, multiparty conversations require interlocutors to be contextualizing cues by evoking frames for inferential understanding. This study revealed that

interlocutors in multiparty discussions utilize lexical cohesive devices to do this. Interlocutors also employ lexical cohesive devices (especially repetition, synonymy, and opposition) for topic continuity strategies. Lexical devices were also used to manage and organize turn-taking behaviors, or to shift focus or drift aspects of global topic. Lexical ties occurred both within and across turns. Therefore, this study has much similar findings with Gonzalez (2010), because here too lexical cohesion is dominant, and the most frequent lexical cohesion are repetition, associative cohesion, and inclusive relations. The two are different in that unlike in multiparty conversations, lexical cohesive devices are not used in telephone conversations as triggers to evoke frames for inferences.

4. Conclusion

This paper explores lexical cohesion as a distinct approach in doing discourse analysis. It runs into four major sections entitled: (1) emergence and nature of cohesion in discourse analysis, (2) scholarly rhetoric and debate on cohesion and textuality in discourse, (3) approaches to lexical cohesion analysis, and (4) insights from lexical cohesion studies. Therefore, in the first two sections, cohesion was defined and discourse-oriented scholars' conflicting views on the phenomenon of cohesion in texts were reviewed. It was shown that while some of these scholars view cohesion as a necessary and sufficient property for text unity, others argue that it is not cohesion but coherence that is the necessary property in texts, and that cohesion is just a byproduct of coherence. It was finally understood that even if cohesion is not considered a necessary property in texts, it remains an essential property in textual communication. It contributes to the achievement and perception of coherence in most texts than not. In the third section, attention was focused on different approaches to lexical cohesion analysis. The study has succeeded in highlighting and discussing the similarities and differences between the models examined. In the last section, different lexical cohesion studies (on written or spoken discourse) were reviewed. In doing this, findings from different studies were carefully compared to show how similar or different they are. It is hoped that this brief review would be of help to students and researchers interested in applying lexical cohesion to research discourse.

References

Angermeyer, P. S. (2002). Lexical cohesion in multilingual conversations. *International journal of Bilingualism*,6(4),

361-393.

Baker, P., &Ellece, S. (2011). *Key terms in discourse analysis*. A&C Black.

Bhatia, V. (2004). *Worlds of written discourse: A genre-based view*. A&C Black.

Blakemore, D. (1992). *Understanding utterances: An introduction to pragmatics* (pp. 155-78). Oxford: Blackwell. Blommaert, J. (2005). *Discourse: A critical introduction*. Cambridge University Press. Bloor, T., & Bloor, M. (2004). *The functional analysis of English*. London: Arnold Bloor, T., & Bloor, M. (2004). *The functional analysis of English*. Routledge.

Brown, G., & Yule, G. (1983). *Discourse analysis*. Cambridge University Press.

Carter, R. (2001). *Working with texts: a core introduction to language analysis*. Psychology Press.

Carter, R., & McCarthy, M. (1988). *Vocabulary and language teaching*. Routledge.

Original Article

Castro, C. D. (2004). Cohesion and the social construction of meaning in the essays of Filipinocollege students writing in L2 English. *Asia Pacific Education Review*, *5*(2), 215- 225.

Creswell, J. W. (2012). *Educafion research: planning, conducting and evaluating quantitative and qualitative research*. Boston/New York: Pearson

De Beaugrande, R. & Dressler, W. (1981). *Introduction to text linguistics*. London: Longman

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford University Press.

Eggins, S. (2004). *Introduction to systemic functional linguistics*. A&C Black. Enkvist, N. E. (1978). *Coherence, Pseudo-Coherence, and Non-Coherence* Flowerdew, J. (2013). *Discourse in English language education*. Routledge.

Gee, J. P. (2011). *How to do discourse analysis: A toolkit*. Routledge.

Gee, J. P. (2011). *An introduction to discourse analysis: Theory and method*. Routledge.

González, M. D. L. Á. G. (2010). Evaluating lexical cohesion in telephone conversations. *Discourse Studies*, 12(5), 599-623.

González, M. D. L. Á. G. (2011). Lexical cohesion in multiparty conversations. *Language Sciences*, 33(1), 167-179.

Gutwinski, W. (1976). *Cohesion in literary texts: a study of some grammatical and lexical features of English discourse* (Vol. 204). Walter de Gruyter.

Halliday, M. A. (1994). *Functional grammar*. London: Edward Arnold.

Halliday, M. A., & Matthiessen, C. M. (2014). *Halliday's introduction to functional grammar*. Routledge.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Routledge.

Hameed, H. T. (2008). Cohesion in Texts: A Discourse Analysis of a News Article in a Magazine. *Al-Faith Journal*. No 37, pp. 81-144

Hasan, R. (1984). Coherence and cohesive harmony. *Understanding reading comprehension: Cognition, language and the structure of prose*, 181-219.

Haswell, R. H. (1988). Critique: Length of text and the measurement of cohesion. *Research in the Teaching of English*, 428-433.

Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Psychology Press.

Hoey, M. (2001). *Textual interaction: An introduction to written discourse analysis*. Psychology Press.

Hoey, M. (1991). *Patterns of lexis in text*.

Original Article

Jabeen, I., Faiz, R., Mehmood, A., & Yousaf, N. (2013). Cohesion and Semantic Understanding. *Academic Research International*, *4*(6), 139.

Jones, R. H. (2012). *Discourse analysis*. London and New York: Routledge.

Kaplan, R. B., &Grabe, W. (2002). A modern history of written discourse analysis. *Journal of second language writing*, 11(3), 191-223.

Khabbazi-Oskouei, L. (2013). Propositional or non-propositional, that is the question: A new approach to analyzing 'interpersonal metadiscourse'in editorials. *Journal of Pragmatics*, *47*(1), 93-107.

Lewin, B., Fine, J., & Young, L. (2001). *Expository discourse*. A&C Black.

Liu, M., & Braine, G. (2005). Cohesive features in argumentative writing produced by Chinese undergraduates. *System*, *33*(4), 623-636.

Lyons, J. (2002). *Language and linguistics*. Cambridge University Press.

Malah, Z. (2015). Lexical Cohesion in Academic Discourse: Exploring Applied Linguistics ResearchArticles Abstracts. *Research Journal of English Language and Literature*,3(4), 291-299.

Malah, Z. (2016). Signals from the Surface of Texts: How Lexical Cohesion Reflects Writers'

Malah, Z., Tan, H., & Rashid, S. M. (2016). Evaluating lexical cohesion in Nigerian newspaper genres: Focus on the editorials. *International Journal of Applied Linguistics and English Literature*, *6*(1), 240-256. Tones in Nigerian Newspaper Editorials. *International Journal Language and Linguistics*, *3*.

Martin, J. R. (2001). Cohesion and Texture. *The Hand Book of Discourse Analysis*.

Martin, J. R. (1992). *English text: System and structure*. John Benjamins Publishing.

Matthiessen, C., Teruya, K., & Lam, M. (2010). *Key terms in systemic functional linguistics*. A&C Black.

McCarthy, M. (1991). *Discourse analysis for language teachers*. Cambridge University Press.

Mirzapour, F., & Ahmadi, M. (2011). Study on lexical cohesion in English and Persian research articles (A Comparative Study). *English Language Teaching*, 4(4), 245.a

Mohamed-Sayidina, A. (2010). Transfer of L1 cohesive devices and transition words into L2 academic texts: The case of Arab students. *RELC Journal*, *41*(3), 253-266.

Morris, J., & Hirst, G. (1991). Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Computational linguistics*, 17(1), 21-48.

Parsons, G. (1990). *Cohesion and Coherence: Scientific Texts* (*A Comparative Study*)

[Monographs in Systemic Linguistics 1]. Nottingham: University of Nottingham, Department of English Studies.

Original Article

Rahman, Z. A. A. A. (2013). The Use of Cohesive Devices in Descriptive Writing by Omani Student-Teachers. *SAGE Open*,*3*(4), 2158244013506715.

Saefudin, S. (2020). Lexical Cohesion Roles In Speech Rhetorical Strategy of 2016 US Presidential Debate, Trump Vs Clinton. *Insaniyat: Journal of Islam and Humanities*, *4*(2), 101-113.

Sanders, T., & Maat, H. P. (2006). Cohesion and coherence: Linguistic approaches. *Reading*, 99, 440-466.

Schiffrin, D., Tannen, D., & Hamilton, H. E. (Eds.). (2008). *The handbook of discourse analysis*. John Wiley & Sons. Sperber, D., Wilson, D. (1986). *Relevance: Communication and cognition* (Vol. 142). Cambridge, MA: Harvard University Press.

Sperber, D., Wilson, D. (2006) in (Eds.). Horn, L., & Ward, G. *The handbook of pragmatics*(Vol. 26). John Wiley & Sons.

Stubbs, M. (2001). Computer-assisted Text and Corpus Analysis: Lexical Cohesion and Communicative Competence. *The handbook of discourse analysis*, *18*, 304.

Taboada, M. T. (2004). *Building coherence and cohesion: Task-oriented dialogue in English and Spanish*(Vol. 129). John Benjamins Publishing.

Tanskanen, S. K. (2006). *Collaborating towards coherence: Lexical cohesion in English discourse*(Vol. 146). John Benjamins Publishing.

Van Dijk, T. A. (2008). *Discourse and context*. *A sociocognitive approach. Cambridge*.

Van Dijk, T. A. (1997). *Discourse as social interaction* (Vol. 2). Sage.

Ventola, E. (1987). *The structure of social interaction: A systemic approach to the semiotics of service encounters*. Pinter Pub Ltd.

Widdowson, H. G. (2004). *Text, context, pretext*. Malden/Oxford: Blackwell Publishing.

Widdowson, H. G. (1978). Teaching language as communication. Oxford University Press.

Wu, S. (2010). Lexical cohesion in oral English. *Journal of Language Teaching and Research*, *1*(1), 97-101.

Yule, G. (2014). *The study of language*. Cambridge University Press.

Yin, Z. (2015). The Use of Cohesive Devices in News Language: Overuse, Underuse or Misuse?. *RELC Journal*, *46*(3), 309-326.