Recent studies in the area of written discourse analysis in English (Batista, 2006; MacMillan, 2006, MacMillan, 2007) suggest that paraphrasing skills, marked by the observance of instances of lexical cohesion in text, play an essential role in the concept of effective reading reflected on the TOEFL® (Test of English as a Foreign Language), a largely accepted standardized English proficiency test designed to determine whether the language skills of international students applying to North American colleges or universities are adequate for enrollment into the selected programs of study.

In reporting on the results of the analysis of 608 fixed-response TOEFL® reading comprehension items using an adapted version of Hoey’s (1991) lexical repetition model, MacMillan (2007) concluded that all question types on the test, in its paper-based, computer-based, and internet-based versions, involve the identification of different forms of lexical repetition connecting question stems and/or correct options to specific sentences in the related passages.

In another application of Hoey’s (1991) repetition model, Károly (1999) demonstrated that a modified version of the same analytical system could be used to evaluate the quality of written discourse structure in argumentative texts produced by advanced EFL learners. Results of her analysis indicated that high-rated and low-rated essays differ significantly in a number of variables related to both the types of cohesive devices employed by learners and the combination of semantic connections between sentences.

The aim of this paper is to propose the use of an adapted version of the lexical cohesive analytical system used in MacMillan (2007) for the practice of academic reading skills in electronically delivered programs. More specifically, this paper introduces e-learning tasks that target the development of paraphrasing skills by means of the identification of different categories of lexical repetition in text.

The following section of this paper will offer a brief outline of the lexical cohesive analytical system used in MacMillan (2007) and its proposed simplified model for use in EAP reading lessons. This will be followed by a detailed description of suggested online reading activities focusing on two basic categories of paraphrasing, namely context-independent and context-bound paraphrasing. The concluding section will discuss means of applying the same strategies to the practice of academic writing skills.

A Theory-based System for the Analysis of Lexical Relations in Text
In his widely acclaimed book, Patterns of Lexis in Text, Hoey (1991) proposed that different forms of lexical repetition combine to organize text. His study has provided evidence that instances of lexical cohesion mark points of reference, or ‘links’, between sentences. The same research has also demonstrated that the observation of repetition patterns in text allows for, among other things, the identification of both adjacent and non-adjacent sentences which have a significant semantic connection.

MacMillan (2007) has devised a modified version of Hoey’s (1991) repetition model and applied it to the identification of bonds between reading comprehension questions and specific portions of the related passages. The resulting taxonomy, represented in Figure 1, below, involves seven types of lexical relations, as well as one set of cohesive devices which are not lexical in nature but which also make it possible for repetition to take place.

The manner in which each of these types of links contributes to the identification of semantic bonds between reading comprehension questions and related passages will now be examined.

The first type of link considered in MacMillan (2007), ‘lexical repetition’, may be classified as either simple or complex. ‘Simple lexical repetition’ (henceforth, ‘simple repetition’) involves the exact repetition of a lexical item (e.g.: drug – drug), or its repetition by means of an inflected form (e.g.: drug – drugs). ‘Complex lexical repetition’ (henceforth ‘complex repetition’), on the other hand, involves the repetition of an item by means of a derived form (e.g.: drug (n.) – drugged (adj)), or by the same form with a different grammatical function (e.g.: drug (n.) – drug (v.)).

The second category of repetition, ‘synonymy’, involves the repetition of the idea represented by a given lexical item, rather than its form. In common with lexical repetition, instances of synonymy may be either simple or complex. ‘Simple synonymy’ occurs...
whenever “a lexical item may substitute for another in context without loss or gain in specificity and with no discernible change in meaning” (Hoey, 1991, p. 62). An example of ‘simple synonymy’ is the repetition of the meaning in the term *able* in one sentence by means of the term *apt* in a different sentence. In its complex form, synonymy involves items which represent equivalent ideas, but which are part of different word classes. An example of ‘complex synonymy’ is the repetition of the meaning in the term *able* in one sentence by means of the term *aptitude* in another.

The third type of lexical relation considered in MacMillan (2007), ‘antonymy’, is also classified as either simple or complex. ‘Simple antonymy’ involves the repetition of the concept of a given item by means of an antonymous term of the same word class. An example of ‘simple antonymy’ is the repetition of concept in the term *violent* in one sentence by means of the contrasting term *peaceable* in another. ‘Complex antonymy,’ on the other hand, involves antonymous terms which are part of different word classes, such as *violent* and *peace*. Note that, following Jones (2002), the term *antonymy* is here used in “its broader sense, referring to any pair of words which could intuitively be recognized as ‘opposites’” (p.1). Therefore, the antonymy category includes not only gradable pairs, such as *cold – hot*, but also non-gradable pairs, such as *dead – alive*, the latter being a category which certain linguists, including Lyons (1977) and Cruse (1986) have termed ‘opposites.’

‘Superordinate’ and ‘hyponymic repetition’ account for cases when two items are interpreted as having identical referents. These links occur when the items sharing the same referent are connected by a lexical relation of class membership. ‘Superordinate repetition’ involves a general term which may be said to designate a class of which the earlier item is a member. One example of ‘superordinate repetition’ is the repetition of the term *painting* in one sentence as the more general term *art* in a subsequent sentence. Conversely, hyponymic repetition involves a specific term which may be said to be a member of, or included in, the class designated by the earlier item forming the link. Therefore, an example of ‘hyponymic repetition’ is the use of the more specific term *painting* to refer back to the term *art* in a previous sentence.

Co-reference links, in common with ‘superordinate’ and ‘hyponymic repetition,’ involve items sharing the same referent. Unlike those, however, co-reference items do not hold a lexical relation, and, thus, the link between them is context-dependent. An example of a co-reference link is the repetition of *Elizabeth II* in one sentence as *the Queen* in a subsequent sentence.

The ‘labeling’ category is based on Francis’ (1994) description of ‘retrospective labels’. The term ‘retrospective label’ refers to a nominal group which encapsulates a stretch of discourse and indicates to the reader how it should be interpreted. The same author (op.cit.) pointed out that these labels are more often than not formed by deictics, such as *this, that or such*, followed by a head noun, which is unspecific in nature, such as Halliday and Hasan’s (1976) ‘general nouns’ (p. 27). Francis (1994) added that a large number of retrospective label head nouns are “metalinguistic in the sense that they label a stretch of discourse as being a particular type of language” (p. 89, original emphasis). An example of a labeling link is the use of the phrase *this question* to refer back to a direct or indirect interrogation within a previous sentence.

Finally, Substitution is the only type of link in this taxonomy which is realized by grammatical members of closed system whose function is to stand in, or substitute for, lexical items. It should be noted that the term *substitution* is here used following Hoey (1991) and Quirk (1972). Most of the items accounted for by this category are described by Halliday and Hasan (1976) as instances of reference. Citing Emmott (1989), Hoey (1991) justified his choice by arguing that “a pronoun, for example, does not refer to an earlier item, but co-refers with the earlier item to something real or imaginary outside the text” (p. 71). However, the differences between Halliday and Hasan’s reference and Hoey’s substitution go beyond the realm of terminology. In addition to personal and demonstrative pronouns, Halliday and Hasan (1976) include demonstrative adjectives, modifiers, and the definite article *the* in their reference category. Because the function of these additional items is largely to draw attention to the givenness of one or more lexical items, rather than to stand in for them, they have not been included in Hoey’s (1991) categorization, or the present study. One exception in this regard is the use of the demonstrative adjectives *this, that, these, and those* to modify a noun which is not a lexical repetition or a paraphrase of a previous item. These instances fall under the labeling category discussed above.

Halliday and Hasan (1976) used the term ‘substitutes’ to refer to a small class of items, some of which are also included in Hoey’s (1991) categorization. One of these items is *one*, when used as a nominal head accompanied by modifiers, as in *the first one* and *another one*. When used by itself, however, *one* would not count as a Substitution link. Rather, it would be treated as accompanying Ellipsis (discussed below).

Other items accounted for in the substitution category as presented in this paper are *do (it/the same/this/likewise/so); the clausal so and not, as in they said so and they said not; and (the) same, when not accompanying an item (repeated or otherwise).*

One final instance of substitution considered in this study is ‘ellipses’, where Θ substitutes for a lexical item. Consider the following example drawn from the sample text in Hoey (1991, p. 227). Sentences are numbered for ease of reference:

[1] It is possible to predict three *reactions* every time a major company, like Barclays Bank, decides to withdraw from the South African economy.
The first option is that the disinvesting firm will insist that its aim is primarily economic and only secondarily political because few businessmen want to admit to yielding to political pressures.

Here, \( \varnothing \) stands in for reaction and acts as the second member of the substitution link formed with reactions, in Sentence 1.

Hoey (1991) argued that substitution items, “while connecting certain sentences, obscure the connections between other sentences” (p.42). Thus, in order to allow for a thorough analysis to be carried out, all sentences in the text must be effectively rendered contextually more neutral. This may be done by replacing non-lexical cohesive features, as well as ellipsis, with the full forms for which they are a shorthand. To exemplify, consider the following sentence (Hoey, 1991, p. 95) followed by its adapted, formatted version (Hoey, 1991, p. 251) with full forms in square brackets:

If it were correct, the writers of political theory would need to be themselves past masters in the art or governing, and statesmen would need to apprentice themselves to them in order to learn their job.

If \(<\text{dit}>\) [the entire conception of politics as an art and of the political philosopher as the teacher of it] were correct, the writers of political theory would need to be themselves past masters in the art or governing, and statesmen would need to apprentice themselves to them in order to learn their job.

MacMillan (2007) has applied the link taxonomy discussed above to the identification of semantic bonds, marked by an above average number of lexical links, between reading comprehension test items, specific sentences in the related passages, and correct options. Consider the following example, based on the introduction to Chapter 2 of Geodesy for the Layman, maintained by the U.S Defense Mapping Agency and available from the National Geodetic Survey web page: www.ngs.noaa.gov.

According to paragraph 2, why may the curvature of the earth be disregarded in plane-table surveys of cities?

- Because the Pythagorean spherical concept is not suitable for simple mathematical calculations.
- Because it does not affect accuracy in determining the relative distance between specific points in small areas.
- Because a flat surface is an acceptable representation of the true figure of the earth in geodetic surveys.
- Because a plane surface provides a more exact figure in astronomical and navigational computations than the sphere does.

(MacMillan, 2007, p. 84)

Each of the options can be joined to the question to form a statement the validity of which can be assessed by means of the identification of a considerable number of links which bond with one or more sentences in the excerpt indicated in the question, namely Paragraph 2. Paragraph 2 is formed by Sentences 6 to 12. The statement formed by the correct option, Option b, bonds with one of the sentences in this excerpt, Sentence 12, by means of as many as nine links, as demonstrated below. Individual links are numbered for ease of reference.

[2] The curvature of the earth may be disregarded in plane-table surveys of cities because it does not affect accuracy in determining the distance between specific points in small areas.

[12] For such small areas, exact positions can be determined relative to each other without considering the size and shape of the total earth.

HOEY’S (1991) repetition model, on which the link taxonomy discussed above is mainly based, was originally devised to identify points of reference between sentences within mainstream, non-narrative texts. Hoey (1991) has established three links as the minimal number of references for two sentences to be considered significantly connected, or bonded. However, this number may vary depending on the relative length and lexical density of the text in question, the cut-off point being marked by an above average degree of repetition cases.

Once links are recorded and bonds between sentences are established, nets, or sets of bonded sentences, can be identified. These nets allow for the identification of marginal and central sentences, as well as topic opening and topic-closing sentences. Marginal sentences are characterized by a very low number of bonds formed with other sentences. Conversely, central sentences are marked by an above average number of bonds formed with other sentences. Finally, topic-opening sentences are characterized by forming most or all of their bonds with subsequent sentences, whereas topic-closing sentences form most or all of their bonds with previous sentences.

One possible means of simplifying this lexical cohesion analytical model for use in reading lessons is by interpreting it as a passive paraphrasing system. The eight different instances of lexical cohesion represented in the link taxonomy may be grouped into two basic categories of paraphrase, namely context-independent and context-bound paraphrase, as shown in Figure 2.
Context-independent paraphrases are marked by instances of lexical cohesion which are more readily observed and only marginally text-bound, i.e., the connection between the lexical links involved is often not limited to the context in which they figure in the text. Instances of repetition of this kind include: Simple and Complex Repetition, Simple Synonymy, Simple Antonymy, as well as Superordinate and Hyponymic Repetition.

Context-bound paraphrases are marked by types of links which are inherently text-bound and which require the identification of either low-level or high-level text-based inference. Low-level text-based inferences involve the identification of referents, which are often represented by instances of Substitution, Ellipsis, Co-Reference, and Labeling. High-level inferences, on the other hand, involve deductions stemming from information implied but not directly stated in the passage. These are often represented by instances of Complex Synonymy and Complex Antonymy.

The following section will demonstrate how this paraphrasing system may be applied to the development of academic reading skills by means of interactive online activities.

**Interactive strategies for the practice of reading skills in online EAP programs**

This section features suggestions to incorporate interactivity into the teaching and practice of passive paraphrasing skills in electronically delivered courses. Suggested content and tasks are presented as forming a full unit in an introductory level EAP program. It is here assumed that learners have been previously guided through the basics of learner-interface interaction.

The unit contains two reading texts, each of which introducing the basic features of one of the paraphrasing modes discussed in Section 2, above. Cognitive activity is encouraged by means of learner-content interaction. Figure 3 includes an example of an activity within the text introducing context-independent paraphrases.

**Look at the following paragraph**

Can you pinpoint in which other sentences the highlighted portion in Sentence 1 reoccurs in the paragraph?

**What forms of repetition are involved?**

1. In technical and encyclopaedic literature one can find somewhat different information about when the word "psychology" was formed and who was the first to use it.
2. In the main psychological and philosophical dictionaries, textbooks, and leading world encyclopaedias there are for the most part three different opinions of the origin of this term which, as the word denoting scientific or philosophic dealing with the phenomena of psychic (subjective, conscious) life, has now come into very wide use.
3. All the three names connected with the formation of the term "psychology" are the names of the people of German origin from the 16th century.
4. Two of them are of little significance: Rudolf Göckel and Otto Casmann, while the third is very famous and generally known: Filip Melanchton.

**Show all reoccurrences**

Here, learners are led to identify clusters of lexical repetition and think about the individual instances of cohesion creating the context-independent paraphrases. Feedback lines appear as learners click on selected portions of the passage. Feedback generated by incorrect choices includes: “Sorry, this is not the correct choice. Try to find excepts that combine alternatives for each of the individual words in the highlighted phrase.” Correct choices, namely “the origin of this term,” in Sentence 2, and “the formation of the term ‘psychology’,” in Sentence 3, generate the following feedback lines, respectively: “Excellent job! Here, when … was formed reoccurs as the origin, and the word psychology as this term;” “Very good! Here, was formed reoccurs as the formation, and the word ‘psychology’ as the term ‘psychology’.” The button “show all reoccurrences” highlights the correct choices and provides the individual instances of cohesion forming the paraphrases as shown in their respective feedback lines.

The reading texts are followed by content review and practice through asynchronous delivery. In this portion of the lesson, knowledge construction is encouraged by means of learner-learner interaction in a discussion board. Figure 4 shows the task used to generate the discussion. The full passage used in this task may be found in the Appendix.

---

1. The passage used in this task consists of an excerpt taken from Kristic’s (1964) Marko Marulic – The Author of the Term “Psychology”.

---
Looking at the following reading comprehension question based on the *Figure of the earth* passage.

Identify the correct option and justify your choice by a) mentioning which sentence(s) in the passage it paraphrases, and b) pointing out the individual semantic connections involved.

Make sure to comment on responses posted by your classmates.

According to paragraph 2, why may the curvature of the earth be disregarded in plane-table surveys of cities?

a. Because the Pythagorean spherical concept is not suitable for simple mathematical calculations.
b. Because it does not affect accuracy in determining the relative distance between specific points in small areas.
c. Because a flat surface is an acceptable representation of the true figure of the earth in geodetic surveys.
d. Because a plane surface provides a more exact figure in astronomical and navigational computations than the sphere does.

In this task, learners are encouraged to identify which sentence(s) in the passage are central to the reading comprehension question, as demonstrated in Section 2, above. The instructor’s role in this discussion is to prompt learners to express whether they agree with responses posted and to offer suggestions to peers. Some feedback may be provided, if necessary. However, in order not to hinder further discussion, the instructor should refrain from identifying the correct option at this point.

The unit is concluded with an online quiz, or short fixed-response reading comprehension test. The first question on the quiz is the same used in the discussion board. When clicking on the correct option, learners are shown a feedback line similar to those used in the learning activity shown in Figure 3, above. The remaining questions should focus on the identification of context-independent paraphrases in order to give learners further opportunities to practice the recognition of less obvious instances of cohesion.

The section that follows will briefly discuss how the paraphrasing strategies exemplified above may be applied to the practice of academic reading skills.

**Applying the Practice of Paraphrasing to the Development of Academic Writing Skills**

The practice of passive paraphrasing skills, as discussed in Section 3, above, may be followed by the practice of active paraphrasing skills, by means of the creation of semantic equivalence using appropriate forms of lexical cohesion.

As a first example, Figure 5 shows how the reading activity in Figure 3 may be modified to target the production of appropriate forms of lexical repetition.

Rewrite the following paragraph replacing unnecessary repetition with appropriate paraphrases.

Make sure to comment on responses posted by classmates.

[1] In technical and encyclopaedic literature one can find somewhat different information about when the word "psychology" was formed and who was the first to use the word “psychology”. [2] In the main psychological and philosophical dictionaries, text books, and leading world encyclopaedias there are for the most part three different opinions of the origin of the word “psychology,” which, as the word denoting scientific or philosophic dealing with the phenomena of psychic (subjective, conscious) life, has now come into very wide use. [3] All the three names connected with the formation of the word “psychology” are names of the people of German origin from the 16th century. [4] Two of the three names connected with the formation of the word “psychology” are of little significance; Rudolf Göckel and Otto Casmann, while the third the name connected with the formation of the word "psychology" is very famous and generally known: Filip Melanchton.

Here, selected items in the original text have been replaced with multiple instances of simple repetition. In this new activity, learners are required to replace redundant forms with appropriate paraphrases and then discuss their choices with their peers. After learners have posted their responses on the discussion board, the instructor may present the original text for comparison. However, students should be encouraged to see the original paraphrases not as ‘the key’ to the exercise, but as examples of possible appropriate semantic connections in that specific context.

Campbell (1990, cited in Jordan, 1997, p. 171) has pointed out that an important aspect of academic writing is “the ability to integrate information from previous researchers in relevant areas of study.” Leki and Carson (1994) have, in turn, recommended that, EAP writing classes encourage learners to integrate personal opinions and experiences with external sources of information. The interactive strategies discussed in this paper may be used to design writing tasks in which learners practice the use of appropriate
paraphrases when integrating information from different sources.

For instance, two short passages discussing opposing views on one specific subject may be used as the basis for essay writing. Directions should lead learners to synthesize information from both passages to support a given argument. Finally, in pairs, students may be encouraged to comment on each other’s use of paraphrase to a) refer to information taken from the reading passages, and b) contribute to the cohesiveness of their essays.

**Conclusion**

This paper has introduced a paraphrasing system based on the identification of different categories of lexical repetition in text. Different interactive e-learning tasks have been proposed as a means of applying paraphrasing strategies to the development of academic reading skills. Finally, suggestions were made as to how to adapt the use of the same strategies to the practice of academic writing skills.

**Appendix - Figure of the Earth**

The expression “figure of the earth” has various meanings in geodesy according to the way it is used and the precision with which the earth’s size and shape is to be defined. The actual topographic surface is most apparent with its variety of land forms and water areas. This is, in fact, the surface on which actual earth measurements are made. It is not suitable, however, for exact mathematical computations because the formulas which would be required to take the irregularities into account would necessitate a prohibitive amount of computations. The topographic surface is generally the concern of topographers and hydrographers.

The Pythagorean spherical concept offers a simple surface which is mathematically easy to deal with. Many astronomical and navigational computations use it as a surface representing the earth. While the sphere is a close approximation of the true figure of the earth and satisfactory for many purposes, to the geodesists interested in the measurement of long distances-spanning continents and oceans-a more exact figure is necessary. The idea of flat earth, however, is still acceptable for surveys of small areas. Plane-table surveys are made for relatively small areas and no account is taken of the curvature of the earth. A survey of a city would likely be computed as though the earth were a plane surface the size of the city. For such small areas, exact positions can be determined relative to each other without considering the size and shape of the total earth.

**References**


