

LEXICAL PATTERNS IN THE READING COMPREHENSION SECTION OF THE TOEFL TEST

Fabiana MACMILLAN¹

■ **ABSTRACT:** The TOEFL (Test of English as a Foreign Language) is currently one of the most widely accepted English language proficiency tests. Designed by the ETS (Educational Testing Service), the main purpose of the TOEFL is to determine whether the English language skills of a student applying to a North American college or university are adequate for enrollment into the selected program of study. This study will focus upon the third section of the TOEFL, Reading Comprehension, which consists of several passages followed by questions with different testing purposes. An adaptation of Hoey's (1991) analytical system for the analysis of lexical cohesion in authentic texts will be used to identify bonds connecting reading comprehension questions on the test to key excerpts in the passages they are related to. A number of sample reading comprehension questions taken from practice tests produced by the ETS will be analyzed. The analysis will focus on the relationship between the testing purpose of each question and the type(s) of lexical link involved in the identification of the correct answer.

■ **KEYWORDS:** Reading Comprehension Tests. Lexical Cohesion. Text Analysis.

¹ Doutoranda do Programa de Pós Graduação em Letras - Instituto de Letras - UFF - 24210-006 - Niteroi-RJ, Brasil. E-mail: fabianam@canadianlineworks.ca

Introduction

The Test of English as Foreign Language (henceforth, TOEFL) is currently one of the most widely accepted English language proficiency tests. Designed by the Educational Testing Service (henceforth, ETS), the main purpose of the TOEFL is to determine whether the English language skills of a student applying to a North American college or university are adequate for enrollment into the selected program of study. In other countries, academic institutions, as well as certain independent organizations, agencies, and foreign governments have also found TOEFL scores useful. In addition, TOEFL scores are also required by a number of medical certification and licensing agencies (*TOEFL Test Score and Data Summary*, 2002-2003 Edition).

The current editions of the TOEFL, namely the computer-based and the paper-based TOEFL, consist of four separately timed sections, namely Listening Comprehension, Structure and Written Expression, Reading Comprehension, and Test of Written English. The aim of this paper is to attempt to identify lexical patterns in the third of these sections, Reading Comprehension. Although it would certainly be of interest to analyze all of the sections of the TOEFL test, it was felt that limitations of time and space would make such an ambitious endeavor fall short on grounds of feasibility.

A group of six Reading Comprehension tests taken from the *TOEFL Test Preparation Kit* produced by the ETS (2002) will be analyzed in terms of the lexical patterns connecting the texts to the questions attached to them. Each of these tests is formed by fifty questions based on five authentic texts in English. The corpus used in this study thus involves 300 reading comprehension questions.

The basic structure of the reading comprehension section is the same in both the paper-based and the computer-based versions of the test. According to the 2000-2001 Edition of the *Computer-based TOEFL Score User Guide*, the reading comprehension section in the computer-based version of the TOEFL consists of a linear test, i.e. "examinees are presented with questions that cover the full range of difficulty (from easy to difficult) as well as the content specifications designated by the test design (p.8)." In this section,

the computer selects for each examinee a combination of passages with accompanying sets of questions that meet both the content

and the statistical designs of the test. The questions are selected without consideration of examinee performance on the previous. The result is a section much like the one on the paper-based test, but each examinee receives a unique set of passages and questions (op.cit, p.8).

The following section of this paper will offer a brief general description of the Reading Comprehension section of the TOEFL as provided by ETS. This description will be followed by an outline of the system of analysis chosen to identify the lexical patterns in each test. The final two sections will discuss the results of the analysis, address its limitations, as well as introduce possible avenues for further research.

The Reading Comprehension Section of the TOEFL

According to the ETS, the Reading Comprehension Section of the TOEFL is designed to measure a student's ability to read and understand short passages in English. The test consists of several passages, each followed by questions. These passages

are about a variety of topics of general interest that are similar to the material that students read in colleges and universities in North America. In other words, the passages reflect fairly formal, objective, neutral, and usually concrete types of academic writing. The subject matter of these passages is general in nature so as not to give an advantage to specialists in particular fields of study, or people with particular kinds of background knowledge. Sufficient context is provided by the passages so that examinees who read and understand them can answer the questions without relying on subject-specific knowledge outside the passage (*TOEFL Test Preparation Kit*, p.56).

The reading comprehension questions about each of the passages serve a number of different testing purposes.² The following table that

² According to the TOEFL official web site (<http://www.ets.org/toefl/>), a new version of the test, the New Generation TOEFL, is to be launched in the United States in September, 2005. New testing purposes and question types will be added to the reading comprehension section in the New Generation TOEFL. In these new questions, test takers may be asked to demonstrate they have learned what they have read in a passage, either by filling out a table or completing a narrative summary. Examinees may also be asked to choose a correct paraphrase of a reading. Because this new version of the TOEFL has not yet been implemented, the new question types it will involve will be disregarded in this study. Therefore, all of the six tests that will be analyzed in this paper have been taken from an ETS practice kit (2002) intended to prepare students for both the paper-based and computer-based editions of the TOEFL test.

comprises each of these testing purposes is as described in the *TOEFL Test Preparation Kit* (p.56). The questions used for exemplification have been taken from different tests in the same publication.

Table 1 Testing Purposes	
1	<p><i>Identify the main topic or the main idea of the passage as a whole or of one of the paragraphs.</i></p> <p>E.g.: Which of the following statements best expresses the main idea of the passage?</p> <ol style="list-style-type: none"> 1. By the nineteenth century, forts were no longer used by the military. 2. Surgeons at forts could not prevent outbreaks of disease. 3. Forts were important to the development of the American West. 4. Life in nineteenth-century forts was very rough.
2	<p><i>Understand some of the details contained in the passage.</i></p> <p>e.g.: Ptarmigan keep warm in the winter by</p> <ul style="list-style-type: none"> • huddling together on the ground with other birds • building nests in trees • burrowing into dense patches of vegetation • digging tunnels into the snow
3	<p><i>Understand the relationships between the ideas in the passage.</i></p> <p>E.g.: What is the relationship between the two paragraphs in the passage?</p> <ol style="list-style-type: none"> 1. The second paragraph explains a term that was mentioned in the first paragraph. 2. Each paragraph describes a different approach to the display of objects in a museum. 3. The second paragraph explains a philosophy of art appreciation that contrasts with the philosophy explained in the first paragraph. 4. Each paragraph describes a different historical period.

4	<p><i>Make inferences based on information that is directly stated.</i></p> <p>E.g.: It can be inferred from the passage that early hotelkeepers in the United States were</p> <ol style="list-style-type: none"> 1. active politicians 2. European immigrants 3. Professional builders 4. Influential citizens
5	<p><i>Identify the meaning of some vocabulary as it is used in the passage.</i></p> <p>E.g.: Look at the word rummage in the passage. Click on another word or phrase in the bold text that is closest in meaning to rummage.</p> <p>Even before humans could make fires themselves, one of the advantages that they (and possibly other primates as well) had over other animals was that they were able to handle sticks with which they could rummage in the smoldering fire without getting burned. After a forest fire they were able to search through the ashes for food and probably noticed that.</p>
6	<p><i>Understand referential relationships in a passage. This includes understanding what word a pronoun refers to as well as understanding what certain abstract concepts (e.g., “this characteristic,” “this idea”) refer to in the context of the passage</i></p> <p>E.g.: The word “those” in line 29 refers to</p> <ul style="list-style-type: none"> • properties • investigations • microscopes • X rays

7	<p><i>Recognize some aspects of how authors organize the information in their writing.</i></p> <p>E.g.: In what order does the author discuss various comic strips in the passage?</p> <ol style="list-style-type: none"> a. In alphabetical order by title b. In the order in which they were created c. According to the newspaper in which they appeared d. From most popular to least popular
8	<p><i>Understand why an author mentions a particular piece of information.</i></p> <p>E.g.: The author mentions the Dakota and the Ansonia in line 24 because</p> <ol style="list-style-type: none"> 1. they are examples of large, well-designed apartment buildings 2. their design is similar to that of row houses 3. they were built on a single building lot 4. they are famous hotels
9	<p><i>Identify the organizational structure of a passage (e.g., compare/contrast, define, chronological sequence).</i></p> <p>E.g.: The author organizes the discussion of forts by</p> <ol style="list-style-type: none"> a. describing their locations b. comparing their sizes c. explaining their damage to the environment d. listing their contributions to western life
10	<p><i>Use the information in the passage to predict how the passage would most likely continue.</i></p> <p>E.g.: The final paragraph of the passage will probably continue with a discussion of</p> <ol style="list-style-type: none"> a. other species of forest birds b. the fragile ecosystem of Newfoundland c. what mammals live in the forests of North America d. how the Newfoundland crossbill survives with a large bill

The next section will demonstrate the system of analysis that will be used to investigate whether lexical patterns underlying the questions realizing each of these testing purposes will vary according to the testing purpose they entail.

Patterns of Lexis in the Reading Section of the TOEFL

Hoey (1991) proposed to describe a new system of text analysis based on the study of cohesion, particularly lexical cohesion. His approach to the study of lexical cohesion is distinguished from that of previous works on cohesion (e.g. HALLIDAY & HASAN, 1976) in that it focuses on observing how cohesive features combine to organize text, rather than on simply identifying and classifying these features. He offers a complete description of a system of analysis for the lexical devices of non-narrative texts, the main features of which will now be briefly described.

Hoey (1991) explains that each sentence in a text contains items that either repeat items from previous sentences or are repeated in the sentences that follow it. He establishes three points of reference as the minimal number of references for two sentences to be considered significantly connected, or *bonded*. However, he adds that in some cases three repetitions, or *links*, may not be sufficient to form a bond between two sentences. Given that the cut-off point of a text is marked by a degree of repetition cases which is above average, it is “related indirectly and uncertainly to the relative length and lexical density of the sentences of the text in question.” (op.cit, p.92)

Hoey (1991, p.51-75) identifies as *links* the kinds of lexical relation which permit repetition, as well as a small set of cohesive devices which are not lexical in nature but which also make it possible for repetition to take place. He has identified nine main types of link, namely [1] simple lexical repetition, [2] complex lexical repetition, [3] simple paraphrase, [4] complex paraphrase, [5] superordinate, [6] hyponymic repetition, [7] co-reference, [8] substitution, and [9] ellipsis. The exemplification of these features will be drawn from an authentic text in English taken from a reading comprehension test in the *TOEFL Sampler CD-Rom* produced by the ETS (2002). All the sentences are numbered for ease of reference

[1] What is it that enabled early humans to control the use of fire; first to keep a fire growing for an extended length of time and then to be successful in passing on this ability from generation to generation? [2] In order to answer this question, it may be useful to distinguish between the physical, mental, and social preconditions that were necessary. [3] No doubt such physical features as erect posture and the concomitant aptitude for carrying objects in the hand and manipulating them were essential. [4] Even before humans could make fires themselves, one of the advantages that they (and possibly other primates as well) had over other animals was that they were able to handle sticks with which they could rummage in the smoldering fire without getting burned. [5] After a forest fire they were able to search through the ashes for food and probably noticed that they might prolong the fire's burning by throwing branches on it. [6] Even more important, however, was the capacity to pick up burning matter and transport it to a place where it could not be extinguished by rain or wind.

[7] But this was clearly not just a matter of the physical advantages of early humans, of erect posture and having the hands free to carry something else. [8] Fetching branches for a fire implies that the individuals concerned thought about what they were doing, and knew why they were doing it. [9] Keeping a fire going implies foresight and care. [10] Wood had to be gathered, and perhaps even stored during wet periods. [11] Such activities did not come naturally to early humans; they required learning and discipline. [12] Especially when humans began to collect fuel over larger distances, they devoted part of their energy to maintaining something outside themselves, something beyond their own immediate needs. [13] This is not to say they were acting "unselfishly." [14] Tending the fire was a form of "deferred gratification" or putting off the satisfaction of immediate needs in planning for future needs, like that which was later to become an essential

ingredient in agriculture and livestock-raising. [15]
Unlike superficially similar complex activities such as nest-building by birds, it was not genetically determined but had to be learned.

According to Hoey (1991, p.53), 'Simple lexical repetition' occurs whenever "a lexical item that has already occurred in a text is repeated with no greater alteration than is entirely explicable in terms of a closed grammatical paradigm." In other words, simple lexical repetition (henceforth 'simple repetition') involves the exact repetition of a content word or its repetition with minimum alterations, such as those marking the plural form of a noun, or those marking the 3rd person singular, simple past, past participle or gerund forms of a verb.

There are a number of examples of simple repetition in the text above, one of which is the following:

Simple Repetition

[2] In order to answer this question, it may be useful to distinguish between the *physical*, mental, and social preconditions that were necessary.

[3] No doubt such *physical* features as erect posture and the concomitant aptitude for carrying objects in the hand and manipulating them were essential.

Complex lexical repetition (henceforth 'complex repetition') occurs "either when two lexical items share a lexical morpheme, but are not formally identical (as defined in our discussion of simple repetition), or when they are formally identical, but have different grammatical functions" (HOEY, 1991, p.55). One example of complex repetition in the text above is the following:

Complex Repetition

[4] Even before humans could make fires themselves, one of the advantages that they (and possibly other primates as well) had over other animals was that they were able to handle sticks with which they could rummage in the smoldering fire without getting *burned*.

[5] After a forest fire they were able to search through the ashes for food and probably noticed that they might prolong the fire's *burning* by throwing branches on it.

Hoey (1991) also argues that paraphrase can also serve the function of repeating. Unlike simple and complex repetition, however, instances of paraphrase repeat the idea represented by a lexical item, not its form. Simple paraphrase occurs whenever “a lexical item may substitute for another in context without loss or gain in specificity and with no discernible change in meaning” (op.cit, p.62). One example of simple paraphrase is:

Simple Paraphrase

[4] Even before humans could make fires themselves, one of the advantages that they (and possibly other primates as well) had over other animals was that they were able to handle sticks with which they could *rummage* in the smoldering fire without getting burned.

[5] After a forest fire they were able to *search through* the ashes for food and probably noticed that they might prolong the fire’s burning by throwing branches on it.

In common with lexical repetition, paraphrase may be either simple or complex. Hoey (1991, p.64) argues that, broadly speaking, complex paraphrase “may be said to occur when two lexical items are definable such that one of the items includes the other, although they share no lexical morpheme.” There are three basic situations to which this definition might be said to apply, namely [1] in cases of antonymous paraphrase, [2] in cases in which a link triangle is identified, and [3] in cases in which a link triangle misses an item. Antonymous paraphrase includes antonymous expressions which do not share a morpheme. E.g.:

Complex Paraphrase

[11] Such activities did not *come naturally* to early humans; they required learning and discipline.

[15] Unlike superficially similar complex activities such as nest-building by birds, it was not genetically determined but *had to be learned*.

The link triangle, Hoey (1991, p.65) explains, occurs when “the presence of two types of link creates a third,” as in the following example:

[1] What is it that enabled early humans to control the use of fire; first to keep a fire growing for an extended length of

time and then to be successful in passing on this *ability* from generation to generation?

[3] No doubt such physical features as erect posture and the concomitant *aptitude* for carrying objects in the hand and manipulating them were essential.

[4] Even before humans could make fires themselves, one of the advantages that they (and possibly other primates as well) had over other animals was that they were *able* to handle sticks with which they could rummage in the smoldering fire without getting burned.

Here, *ability* and *aptitude* form a simple paraphrase link, and *ability* and *able* form a complex repetition link. Therefore, *aptitude* and *able* will form a complex paraphrase. However, Hoey (1991) adds, there will be situations in which one of the elements of the triangle is missing. He argues that a link may be acknowledged between the two items present in the text if certain conditions are met: "there must be an item that is capable of paraphrasing exactly in that context one of the items and of repeating the other. In other words, the missing item has to be such that if it were to be substituted for the item it paraphrases there would be no discernible difference in our interpretation of the text." (op.cit, p.66) Thus, if, for instance, a text contains the words *teacher* and *instruction*, these may be considered to form a link, given that the missing item, *teaching*, can substitute exactly for *instruction* in this context.

Hyponymic repetition, as well as Superordinate and Co-reference, account for cases when two items are interpreted as having identical referents. Superordinate and hyponymic repetition occur when the items sharing the same referent are lexically related. Superordinate refers to a more general term, such as *primates* (Sentence 4), whereas hyponym refers to a more specific term, *humans* (Sentence 4). Co-reference, on the other hand, occurs when the two items sharing the same referent are not lexically related. Two examples of co-reference links in the text above are the following:

Co-reference

[1] *What is it that enabled early humans to control the use of fire; first to keep a fire growing for an extended length of time and then to be successful in passing on this ability from generation to generation?*

[2] In order to answer *this question*, it may be useful to distinguish between the physical, mental, and social preconditions that were necessary.

[1] What is it that enabled *early humans* to control the use of fire; first to keep a fire growing for an extended length of time and then to be successful in passing on this ability from generation to generation?

[8] Fetching branches for a fire implies that *the individuals concerned* thought about what they were doing, and knew why they were doing it.

The eighth type of link, Substitution, is realized by textual, rather than lexical items. These items are grammatical members of closed systems whose function is to stand in, or substitute for, lexical items. In Hoey's (1991) categorization, substitutes include personal pronouns, demonstrative pronouns and modifiers. The same category also includes *one* when used as a nominal head and modified in some way. This can be seen in *the first one* and *another one*; *do* in (*do it/the same/this/likewise/so*); the clausal *so*, as in *they said so*; and the demonstrative modifiers (*an*)*other*, (*the*) *other*, (*the*) *same*, *different* and *similar*. One example of a Substitution link in the text above is the following:

Substitution

[15] Unlike superficially similar complex activities such as nest-building by birds, *<it>* [*tending the fire* – sentence 14] was not genetically determined but had to be learned.

Finally, Ellipsis is described by Hoey (1991, p.74) as “the absence of some required stretch of language that has to be supplied by the listener or reader to make sense of the sentence encountered.” E.g.:

Ellipsis

[2] In order to answer this question, it may be useful to distinguish between the physical, mental, and social preconditions that were necessary [*for early humans to be able to control fire*].

Hoey (1991, p.74) adds that “ellipsis has to be accounted for whether or not it gives rise to grammatical incompleteness, when we

consider the ways in which we may produce sub-texts from the non-narrative texts we analyse.”

These examples for Substitution and Ellipsis demonstrate that certain links are realized by non-lexical cohesive features including pronouns and determiners. In terms of these items, Hoey (1991, p.42) argues that: “while connecting certain sentences, obscure the connections between other sentences.” 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, an adapted/formatted version of an excerpt of the text above may appear as follows (full forms are in square brackets):

[5] After a forest fire <they> [early humans] were able to search through the ashes for food and probably noticed that they might prolong the fire’s burning by throwing branches on it.

[6] Even more important [than the ability to search through the ashes for food and notice they might prolong the fire’s burning by throwing branches on it – sentence 5], however, was the capacity to pick up burning matter and transport it to a place where it could not be extinguished by rain or wind.

Hoey’s (1991) system was originally devised to identify the links bonding both adjacent and non-adjacent sentences within mainstream non-narrative texts. However, the same system may be applied to identify bonds between reading comprehension questions and the sentences in the text containing the answers to them (BATISTA, 2002). To exemplify, consider the following reading comprehension questions about the text above, also taken from the *TOEFL Sampler* CD-Rom produced by the ETS (2002):

In paragraph 2, the author suggests that the controlling of fire by early humans

- a. was based on instinct
- b. was made difficult by their posture
- c. required the ability to plan ahead
- d. was ‘unselfish’ behavior

This question type realizes the second testing purpose presented in Section 2 above, namely “to understand some of the details contained in the passage.” 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 bonding it to one or more sentences in the excerpt indicated in the question, namely paragraph 2, which is formed by Sentences 7 to 15. The statement formed by option *c* above is that which contains the largest number of links connecting it to sentences in this excerpt, as demonstrated below:

The author suggests^A that the controlling of fire^B by early humans^C required the ability to plan ahead^D.

[8] Fetching branches for a fire^B implies^A that the individuals concerned^C thought about what they were doing, and knew why they were doing it.

[9] Keeping a fire going^B implies^A [that the individuals concerned demonstrated] foresight^D and care.

The statement formed by option *c* is bonded to sentence 8 in the text by means of four links, namely (A) *suggests – implies* (simple paraphrase), (B) *controlling of fire – fetching branches for a fire* (superordinate), (C) *early humans – the individuals concerned* (co-reference), and (D) *plan – thought about* (simple paraphrase). The statement is also bonded to sentence 9 through the four following links: (A) *suggests – implies* (simple paraphrase), (B) *controlling of fire – keeping a fire going* (superordinate), (C) *early humans – Ø* (ellipsis of *the individuals concerned*), and (D) *plan ahead – foresight* (complex paraphrase).

In the example above, a bond connecting the question (with its embedded statement formed by the right option) to the assigned excerpt in the text was marked by a number of different kinds of links, the most frequent of which in this case is paraphrase. However, other questions will involve the identification of a single kind of link, as in the following selected examples³:

3. Look at the word them in the passage. Click on the word or phrase in the bold text that them refers to.

No doubt such physical features as erect posture and the concomitant aptitude for carrying objects in the hand and manipulating them were essential.

4. Look at the word **rummage** in the passage. Click on another word or phrase in the bold text that is closest in meaning to **rummage**.

Even before humans could make fires themselves, one of the advantages that they (and possibly other primates as well) had over other animals was that they were able to handle sticks with which they could **rummage** in the smoldering fire without getting burned. After a forest fire they were able to search through the ashes for food and probably noticed that they might prolong the fire's burning by throwing branches on it.

5. Look at the word **it** in the passage. Click on the word or phrase in the bold text that **it** refers to.

Tending the fire was a form of "deferred gratification" or putting off the satisfaction of immediate needs in planning for future needs, like that which was later to become an essential ingredient in agriculture and livestock-raising. Unlike superficially similar complex activities such as nest-building by birds, **it** was not genetically determined but had to be learned.

Questions 3 and 5 realize the sixth testing purpose introduced in the previous section, namely "understand the referential relationships in a passage." In both questions the student is asked to identify a substitution link, the former within the same sentence (*objects – them*), and the latter within adjacent sentences (*tending the fire – it*). In Question 4, realizing the fifth testing purpose introduced in Section 2 above, the student is asked to identify a simple paraphrase link (*rummage – search though*).

³ Questions 1 and 2 the test referred to were not considered relevant to purpose of this research.

Questions which require the student to add a given sentence to the text, as in Question 6 below, can be associated with the third testing purpose, discussed previously, namely “understand relationships between ideas in the passage.” Question 6 involves the identification of a bond connecting a suggested sentence with one of eight option sentences marked by squares in the text:

6. The following sentence can be added to Paragraph 2.

On the contrary, in caring for the fire they were also caring for themselves.

Where would it best fit in the paragraph? Click on the square (■) to add the sentence to the passage (Sentences 7 to 15 in example text printed on page 7).

Sentence 13 is the one which contains the largest number of links bonding it to the sentence to be added, as demonstrated below:

[13] <This> [The fact that early humans^A would collect fuel over larger distances^B and devote part of their energy to maintaining something outside themselves^C, something beyond their immediate needs – sentence 12] is not to say^D <they> [early humans] were acting “unselfishly.”

On the contrary^D, in caring for the fire^B they^A were also caring for themselves^C.

The sentence which is supposed to be added to the text supports the idea presented in Sentence 13 by explaining why one should not assume early humans were acting unselfishly by devoting part of their energy to maintain something which was beyond their immediate needs. The main kind of link bonding these sentences is the paraphrase. Simple paraphrase was used to repeat the ideas in Sentence 13 marking the theme, or point of departure (HALLIDAY, 1994), of the sentence to be added (*this is not to say – on the contrary*). Complex (antonymous) paraphrase (*maintaining something outside themselves – caring for themselves*) was used to highlight the rheme, or the new information (LOCK, 1996), of the sentence to be added. The apparently opposing

ideas emphasize the goal-achievement relation between both sentences. Rather than being their ultimate goal, caring for the fire was actually the means early humans found to achieve their actual goal: Caring for themselves.

One last question type in the test used here for exemplification, Question 1 below, requires students to identify the main idea in the text, and thus realizes the first testing purpose as described by the ETS (2002, p.56):

1. Which of the following is the main topic of the passage?
 - a. The positive effects of forest fires on early humans.
 - b. Early indications of superior human intelligence.
 - c. Characteristics that made it possible for early humans to control fire.
 - d. Environmental conditions that threatened the survival of early humans.

This kind of question, Enright et al. (2000, p.14) argue, “is likely to require cycling through and integrating a range of information from various points in the text.” They add that tasks that are intended to test basic comprehension might require examinees to distinguish main ideas from minor ideas or inferring the main topic. It follows that questions of this type are likely to require the identification of a number of different links connecting the right option to different points in the text.

The first few sentences in the text above, for instance, might lead one to suppose the right answer to Question 1 is option *c*, namely “characteristics that made it possible for early humans to control fire.” However, a more careful analysis of the text as a whole shows that there is a larger number of direct and indirect references in the text to the idea put across in option *b*, “early indications of superior human intelligence.” The five content words in option *b* form links with all but two sentences in the text, as Table 2 demonstrates. The first item in each pair is from option *b*; the second item is from the numbered sentence.

Table 2 - Lexical Links in “Early Humans” text

Sentence #	Links
1	Simple repetition: early – early Complex repetition: human – humans
2	Ellipsis: early – Ø Ellipsis: human – Ø Complex paraphrase: intelligence – mental
3	
4	Complex repetition: human – humans Complex paraphrase: superior – advantages
5	Substitution: early – they Substitution: human – they Complex paraphrase: intelligence – noticed
6	Ellipsis: intelligence – Ø
7	Simple repetition: early – early Complex repetition: human – humans Complex paraphrase: superior – advantages
8	Complex paraphrase: indications – implies Complex paraphrase: intelligence – thought
9	Complex paraphrase: indications – implies Complex paraphrase: intelligence – foresight
10	
11	Simple repetition: early – early Complex repetition: human – humans Complex paraphrase: intelligence – learning
12	Complex repetition: human – humans
13	Substitution: early – they Substitution: human – they
14	Complex paraphrase: intelligence – planning
15	Ellipsis: early – Ø Ellipsis: human – Ø Complex paraphrase: superior – complex Complex paraphrase: intelligence – learned

Table 2 shows that in Question 1, as occurred in questions 2 and 6, paraphrase predominates and is, therefore, the main kind of link used to connect the right option to key excerpts in the text. Note that questions 1, 2 and 6 in this test are precisely the ones that do not involve a single specific kind of link, as occurred in questions 3, 4 and 5, involving substitution, paraphrase, and substitution links respectively. This fact suggests that there might be a lexical pattern underlying each type of question in the reading comprehension section of the TOEFL. Analogous to this hypothesis is the possibility of associating reading comprehension tests in the TOEFL to Hoey's (2001) definition of a discourse colony.

According to Hoey (2001, p.75) "a colony is a discourse whose component parts do not derive their meaning from the sequence in which they are placed." In other words, unlike mainstream or continuous prose texts, the parts of a discourse colony may be jumbled without affecting overall meaning. Among the numerous text types which conform to this definition, Hoey (2001) mentions exam papers, criminal statutes, and telephone directories. He explains that the crucial point associating these discourse types to the working definition of a colony is that in none of them are the component parts dependent on their neighbors for meaning. To exemplify, Hoey (op.cit, p.75) mentions that "the function and meaning in a telephone directory is obtained from its place in the whole, not its place in a sequence." In the case of reading comprehension tests, the questions may be jumbled without affecting the function or meaning of the tests as a whole.

Hoey (2001) further suggests that the component parts to discourse colonies are connected by lexical cohesion. He notes, for instance, that the sections in criminal statutes are frequently, and almost exclusively, connected by heavy lexical repetition. The brief analysis of the test above suggests that reading comprehension tests within the TOEFL similarly derive their texture, or unity with respect to their environment, by means of lexical cohesion (HALLIDAY & HASAN, p.1976). As previously mentioned, the same analysis suggests that lexical patterns may be identified according to question type. In order to check this hypothesis, a group of six reading comprehension tests taken from the *TOEFL Test Preparation Kit* produced by the ETS were analyzed in terms of the lexical patterns connecting the texts to the questions attached to them. The following section will briefly discuss the results of the analysis.

Analysis and Discussion

The analysis of the six reading comprehension tests in the *TOEFL Preparation Kit* suggests that two testing purposes, 5 and 6, almost invariably involve the identification of a single type of lexical link. The fifth testing purpose presented by the ETS, "identify the meaning of vocabulary," accounts for about 30% of the questions in the corpus. Most of these questions require that the student identify a simple paraphrase link, connecting a given word in the passage, with one of the words given as options, as in the following example. The location of the questions that are used for exemplification hereafter is given in brackets. The letter refers to the test from which the question was taken: A, B, C, D, E, or F. The first number, within hyphens, refers to the text to which the question is attached: 1, 2, 3, 4, or 5. The last two numbers, 01 to 50, refer to the number of the question within that test. Whenever necessary, formatted versions (as demonstrated in Section 3 above) of the sentences referred to in the questions are given in italics.

(A-1-03)

The word "dictates" in line 8 is closest in meaning to

- a. reads aloud
- b. **determines**
- c. includes
- d. records

*For example, sometimes the shape or veining in a piece of stone or wood suggests, perhaps even **dictates**, not only the ultimate form, but even the subject matter.*

A small number of questions, however, will require the student to identify the meaning of a clause rather than a word or phrase, as in the following example:

(C-2-14)

What does the author mean by stating that "The **dulotic species of ants^A** . . . are the supreme social **parasites^B**" (line 5)?

- a. The Polyergus are more highly developed than the Formica.
- b. The Formica have developed specialized roles.
- c. **The Polyergus^A are heavily dependent^B** on the Formica.

d. The Formica do not reproduce rapidly enough to care for themselves.

In this case, the correct option is linked to the selected excerpt of the passage by a co-reference link (*The dulotic species of ants – The Polyergus*), in addition to the paraphrase link characteristic of the question's testing purpose, here between *parasites* and *heavily dependent*.

The sixth testing purpose, "understand the referential relationships in the passage," accounts for nearly 10% of the questions in the corpus. Almost all of the questions in the corpus, realizing this testing purpose, involve the identification of a substitution link connecting one of the options to a given sentence in the passage, as in the following:

(F-1-02)

The word "they" in line 5 refers to

- a. track's
- b. trains
- c. freight, mail, and passengers
- d. steamboats, wagons, stagecoats

This meant that wagon freighting, stagecoaching, and steamboating did not come to an end when the first train appeared: rather they became supplements or feeders.

One example of a question realizing the same testing purpose in this corpus, however, involves the identification of a co-reference link, rather than the usual substitution link.

(B-1-01)

The word "medium" in line 5 could be used to refer to

- a. stone or wood
- b. mallet or chisel
- c. technique
- d. principle

Implicit in it> [the direct carving technique] is an aesthetic principle as well: that the medium> [the stone or wood used by carvers] has

certain qualities of beauty and expressiveness with which sculptors must bring their own aesthetic sensibilities into harmony.

All of the other testing purposes realized by the questions in this corpus involve the identification of a number of different types of link, connecting a given excerpt of the passage, to the right option. However, the analysis suggests that certain kinds of link are more frequently used, or are central to the identification of the correct option, in questions realizing given testing purposes. For instance, paraphrase links seem to predominate in questions realizing the first and second testing purposes described by the ETS, namely, “identify the main topic or the main idea of the passage as a whole or of one of the paragraphs,” and “understand some details contained in the passage,” respectively.

Hoey (1991, p.113) claims that sentences having a level of bonding (usually marked by a minimum of three links) which is above average might be regarded as central to the development of the theme(s) of a text. However, reading comprehension questions requiring examinees to identify the main theme of a passage or paragraph sometimes offer options, which involve less than two content words. Thus, the correct option in questions realizing this testing purpose would mainly involve words, which individually connect with other words in the largest number of sentences in the passage. Consider the following example in this corpus:

(F-2-11)

What does the passage mainly discuss?

- a. The efforts of early humans to care for herds of animals
- b. The development of writing
- c. The beginnings of mathematics
- d. Similarities in number sense between humans and animals

All of the options contain words that link with at least one sentence in the passage the question is attached to. The two content words (beginnings, mathematics) in option *c*, however, independently link with words in all of the 11 sentences in the passage, as demonstrated by Table 3 below.

Table 3 - Lexical Links in Questions F-2-11

Sentence #	Links
1	Complex paraphrase: beginnings – earliest Simple repetition: mathematics – mathematics
2	Complex paraphrase: beginnings – developed Complex paraphrase: mathematics – counting
3	Complex paraphrase: beginnings – came about Substitution: mathematics – it
4	Complex paraphrase: beginnings – prehistoric Complex paraphrase: mathematics – number sense
5	Simple paraphrase: beginnings – evolution Complex paraphrase: mathematics – counting
6	Complex paraphrase: beginnings – earliest Complex paraphrase: mathematics – count
7	Complex paraphrase: mathematics – count
8	Complex paraphrase: mathematics – counts
9	Complex paraphrase: beginnings – later Complex paraphrase: mathematics – number
10	Complex paraphrase: beginnings – later Complex paraphrase: mathematics – numbers
11	Complex paraphrase: beginnings – development

Note that all but two of the links, which the words in the right option form with terms in the passage, are paraphrase links. Paraphrase links are equally frequent in questions whose purpose is to test an examinee’s ability to understand details in the passage, as in the example below:

(D-2-14)

According to the author, the steel wires used to make barbed wire are specially processed to

- a. protect them against rust
- b. make them more flexible
- c. prevent contraction in cold weather
- d. straighten them

As previously discussed in Section 2, questions realizing this testing purpose usually involve the combination of the wording in the question to that in the right option, in order to form a statement which should connect with an excerpt in the passage by an above average number of links. The statement formed by the right option in this question, option *a*, bonds with the eighth sentence in the passage, as demonstrated below:

The steel wires^A used to make barbed wire^B are specially processed^C to protect them against rust^D.

The steel wires^A used [to make modern barbed wire^B] are galvanized^C — coated with zinc to make them rustproof^D.

The element in the statement corresponding to option *a*, “protect them against rust” forms a simple paraphrase link with “make them rustproof” in sentence 8.

In a smaller number of cases in this corpus, however, questions realizing Testing Purpose 2, involve the identification of repetition links rather than paraphrase, as in the following:

(D-1-08)

According to the passage, what made it almost impossible for other groups to conquer the Anasazi?

- a. The political and social organization of the Anasazi
- b. The military tactics employed by the Anasazi
- c. The Anasazi’s agricultural technology
- d. The natural barriers surrounding Anasazi villages

Here, the statement formed by the correct option, option *a*, bonds with the last sentence in the passage by means of simple repetition links:

The political and social organization of the Anasazi made it almost impossible for other groups to conquer the Anasazi.

The cohesive and political social organization of the Anasazi made it almost impossible for other groups to conquer them.

Note that these two questions have different levels of difficulty.

In the review section of the *TOEFL Practice Kit* (p.336), the former question is rated as medium difficulty, whereas the latter is rated as easy. This fact suggests that the level of difficulty of a question might be marked by the type of lexical link required to successfully answer it. Thus, the identification of a paraphrase link would represent a higher level of difficulty; the identification of a repetition link, a lower level of difficulty.

Very few questions clearly representing testing purposes 3, 7, 9 and 10 were found in this corpus. Questions realizing these testing purposes, when combined, account for no more than 1% of the total questions in the corpus. Therefore, the lexical links connecting these questions with excerpts in the passages they refer to could not be considered as representative of a pattern.

Questions realizing the fourth testing purpose, “make inferences based on information directly stated,” account for around 8% of the questions in the corpus. In common with those reflecting testing purposes 1 and 2, these questions tend to mainly require the identification of both repetition and paraphrase links. Thus, the following example might be considered:

(B-3-24)

It can be inferred that railroad refrigerator cars came into use

- a. before 1860
- b. before 1890
- c. after 1900
- d. after 1920

The term ‘railroad refrigerator cars’ figures in only of the sentences in the passage. This sentence is immediately followed by a sentence that mentions the 1890’s. A simple paraphrase link between the words *before* and *by* identifies option *b* as the correct option as demonstrated below:

Railroad refrigerator cars^A came into use before^B 1890^C.

Railroad refrigerator cars^A enabled growers and meat packers to ship perishables great distances and to preserve them for longer periods. Thus, by^B the 1890’s^C, northern city dwellers could enjoy southern and western strawberries, grapes, and tomatoes, previously available for a month at most, for up to six months a year.

Questions realizing the eighth testing purpose, “understand why an author mentions a particular piece of information,” concern the function of a given statement in relation to its adjacent context; in this corpus, this would involve the one or two sentences that precede the sentence in question. Thus, in a number of cases, the options offered will involve such verbs as “describe”, “compare”, “explain”, “illustrate” and “emphasize”, as in the following question:

(E-5-49)

Why does the author mention the Virgo galaxy and the Andromeda galaxy in the third paragraph?

- a. To describe the effect that distance has on visibility
- b. To compare the ages of two relatively young galaxies
- c. To emphasize the vast distances of the galaxies from Earth
- d. To explain why certain galaxies cannot be seen by a telescope

The author of the passage in question mentions the Andromeda galaxy in the third sentence, in the third paragraph, and the Virgo galaxy in the last sentence in the same paragraph. These sentences, here transcribed in their formatted version, together with the one sentence that immediately precedes each, bond with the statement formed by option *c* as demonstrated below:

The author mentions the Virgo galaxy^A and the Andromeda galaxy^B to emphasize^C the vast^D distances^E of the galaxies^F from Earth^G.

By comparison with <these familiar yardsticks> [terrestrial distances expressed as intervals of time], the distances^E to the galaxies^F are incomprehensibly^C large^D, but they too are made more manageable by using a time calibration, in this case, the distance that light travels in one year.

On <such a scale> [one light year] the nearest^C giant spiral galaxy, the Andromeda galaxy^B, is two million years away^E.

<Their> [of the most^C distant^E luminous objects seen by

*telescopes/ light was already halfway here before the **Earth^a** even formed.*

*The light from the **nearby^e** > [from **Earth^a**] **Virgo galaxy^a** set out when reptiles **still^f** dominated the animal world.*

Note that evaluative words denoting emphasis, marked with the letter C, were here considered to form a complex paraphrase link with the verb “emphasize” in option c. Other paraphrase links identifying option c as the correct answer to the question are *distances – away*, and *distances – nearby*.

A larger number of questions realizing the eighth testing purpose in this corpus, however, concern statements in the passage that serve as an exemplification of an idea put forward in previous sentences, as in the following:

(B-2-13)

The author mentions kinglets in line 9 as an example of birds that

- a. protect themselves by nesting in holes
- b. nest with other species of birds
- c. nest together for warmth
- d. usually feed and nest in pairs

The author mentions kinglets in the last sentence of the first paragraph in the passage. This sentence, as well as the sentence that immediately precedes it, bond with the statement formed by option c, as demonstrated below:

***Kinglets^a** are an example of **birds^b** that **nest^c together^d** for **warmth^e**.*

*Body **contact^d** **reduces^f** the surface area exposed to the cold air, so the **birds^b** keep each other **warm^e**.*

*Two **kinglets^a** **huddling^c together^d** were found to **reduce^f** their **heat^e** losses by a quarter, and three together saved a third of their heat.*

Once again, paraphrase links predominate.

Very few questions clearly representing testing purposes 3, 7, 9 and 10 were found in the tests analyzed in this study. Questions realizing these testing purposes, when combined, account for no more than 1% of the questions in the corpus. Therefore, the lexical links connecting these questions with excerpts in the passages they refer to could not be considered as representative of a pattern.

Conclusion

The analysis of the six reading comprehension practice tests in the *TOEFL Preparation Kit* seems to confirm the hypothesis that given testing purposes involve the identification of specific types of lexical links. The analysis in certain cases also suggests that the difference in the type of link, connecting the right option to key excerpts within the passage in question, accounts for the level of difficulty of a question. However, the results reached in this study may be considered as no more than initial, given that the number of questions realizing certain testing purposes in the corpus proved to be unrepresentative.

Clearly, further research involving larger corpora is needed to provide further evidence that lexical patterns underlying reading comprehension questions in the TOEFL can be identified. Care should be taken, however, that such corpora include a representative number of questions realizing all of the basic testing purposes in the Reading Comprehension Section on the TOEFL. More importantly, it is necessary to investigate the extent to which EFL learners might benefit from the explicit teaching of these patterns.

Padrões lexicais na seção de leitura e compreensão do TOEFL

■ **RESUMO:** *O TOEFL (sigla em inglês para Teste de Inglês como Língua Estrangeira) é atualmente um dos mais amplamente aceitos testes de proficiência em inglês ao redor do mundo. Elaborado pela ETS (sigla em inglês para Serviço de Testes Educacionais), o principal objetivo do TOEFL é determinar se as habilidades lingüísticas em inglês de*

um candidato estrangeiro a universidades e faculdades norte-americanas são adequadas para sua admissão ao curso pretendido. O objeto deste estudo é a terceira seção do teste, Leitura e Compreensão de Textos, que consiste em várias passagens em inglês seguidas de perguntas com diferentes propósitos avaliativos. Uma adaptação do sistema de análise de elementos coesivo-lexicais proposto por Hoey (1991) será usada com o objetivo de identificar laços coesivos unindo questões de leitura e compreensão a pontos-chave nas passagens a que se referem. Questões de leitura e compreensão retiradas de testes práticos para o TOEFL produzidos pela ETS serão analisadas especificamente quanto à relação entre seus propósitos avaliativos e os tipos de laços coesivos envolvidos na identificação das respostas corretas.

■ **PALAVRAS-CHAVE:** *Testes de Leitura e Compreensão. Coesão Lexical. Análise do Texto.*

References

BATISTA, F.M. *Investigating Reading Strategies Using Think Aloud Protocols*. Dissertação (Mestrado em Letras) – Instituto de Letras, Universidade Federal Fluminense. Rio de Janeiro, 2002.

Computer-Based TOEFL Score User Guide. ETS, 2000-2001 Edition.

ENRIGHT, M.K. et al. *TOEFL 2000 Reading Framework: A Working Paper*. TOEFL Monograph Series, April 2000.

HALLIDAY, M.A.K. *An Introduction to Functional Grammar*. (2nd ed). London: Edward Arnold, 1994.

HALLIDAY, M.A.K and HASAN, R. *Cohesion in English*. London: Longman, 1976.

HOEY, M. *Patterns of Lexis in Text*. Oxford: OUP, 1991.

_____. Textual Interaction. London: Routledge, 2001.

LOCK, G. Functional English Grammar. Cambridge: CUP, 1996.

TOEFL Test Preparation Kit and TOEFL Sampler CD-Rom. ETS, 2002.

TOEFL Test Score and Data Summary. ETS, 2002-2003 Edition.