Coherence, Cohesion, and Writing Quality

Stephen P. Witte; Lester Faigley

College Composition and Communication, Vol. 32, No. 2, Language Studies and Composing. (May, 1981), pp. 189-204.

Stable URL:
http://links.jstor.org/sici?sici=0010-096X%28198105%2932%3A2%3C189%3ACCAWQ%3E2.0.CO%3B2-E

College Composition and Communication is currently published by National Council of Teachers of English.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/ncte.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.

http://www.jstor.org
Sat Jan 26 11:13:05 2008

Coherence, Cohesion, and Writing Quality

Stephen P. Witte and Lester Faigley

A question of continuing interest to researchers in writing is what internal characteristics distinguish essays ranked high and low in overall quality. Empirical research at the college level has for the most part taken two approaches to this question, examining errors[1] and syntactic features[2] while generally ignoring the features of texts that extend across sentence boundaries.[3] Neither the error approach nor the syntactic approach has been entirely satisfactory. For example, Elaine Maimon and Barbara Nodine's sentence-combining experiment suggests that, as is true when other skills and processes are learned, certain kinds of errors accompany certain stages in learning to write.[4] Because the sources of error in written discourse are often complex and difficult to trace, researchers can conclude little more than what is obvious: low-rated papers usually contain far more errors than high-rated papers. With regard to syntax, Ann Gebhard found that with few exceptions the syntactic features of high- and low-rated essays written by college students are not clearly differentiated. Indeed, research in writing quality based on conventions of written English and on theories of syntax, particularly transformational grammar, has not provided specific directions for the teaching of writing.

Such results come as no surprise in light of much current research in written discourse. This research—published in such fields as linguistics, cybernetics, anthropology, psychology, and artificial intelligence—addresses questions, concerned with extended discourse rather than with individual sentences, questions about how humans produce and understand discourse units often referred to as *texts.*[5] One such effort that has attracted the attention of researchers in writing is M. A. K. Halliday and Ruqaiya Hasan's *Cohesion in English*.[6] Although Halliday and Hasan do not propose a theory of text structure or examine how humans produce texts, they do attempt to define the concept of *text.* To them a text is a semantic unit, the parts of

Stephen Witte and Lester Faigley teach at the University of Texas at Austin. Both have published essays on research in the teaching of writing. They serve as co-directors of a research project on the evaluation of programs and courses in writing, funded by the Fund for the Improvement of Post-Secondary Education.

which are linked together by explicit cohesive ties. Cohesion, therefore, defines a text as text. A *cohesive tie* "is a semantic relation between an element in a text and some other element that is crucial to the interpretation of it" (p. 8). The two semantically connected elements can lie within the text or one element can lie outside the text. Halliday and Hasan call within-text cohesive ties *endophoric* and references to items outside the text *exophoric.* An example of an exophoric reference is the editorial "we" in a newspaper. Such references are exophoric because no antecedent is recoverable within the text. Exophoric references often help link a text to its situational context; but, as far as Halliday and Hasan are concerned, exophoric references do not contribute to the cohesion of a text. For Halliday and Hasan, cohesion depends upon lexical and grammatical relationships that allow sentence sequences to be understood as connected discourse rather than as autonomous sentences. Even though within-sentence cohesive ties do occur, the cohesive ties across "sentence boundaries" are those which allow sequences of sentences to be understood as a text.

Halliday and Hasan's concept of textuality, defined with reference to relationships that obtain across "sentence boundaries," suggests a number of possibilities for extending composition research beyond its frequent moorings in sentence-level operations and features. The major purpose of the present study is to apply two taxonomies of cohesive ties developed by Halliday and Hasan to an analysis of essays of college freshmen rated high and low in quality. Because *Cohesion in English* is a pioneering effort to describe relationships between and among sentences in text, we anticipate that cohesion will be studied in future research addressing the linguistic features of written texts. We are particularly interested in identifying what purposes Halliday and Hasan's taxonomies can serve in composition research and what purposes they cannot serve.

Halliday and Hasan's System for Analyzing and Classifying Cohesive Ties

Cohesion in English specifies five major classes of cohesive ties, nineteen subclasses, and numerous sub-subclasses. In the analysis of cohesion which follows, we will be concerned with only the five major classes—*reference, substitution, ellipsis, conjunction,* and *lexical reiteration and collocation*—and their respective subclasses. Two of the major classes—*substitution* and *ellipsis*—are more frequent in conversation than in written discourse. *Substitution* replaces one element with another which is not a personal pronoun, and *ellipsis* involves a deletion of a word, phrase, or clause. The effect of both substitution and ellipsis is to extend the textual or semantic domain of one sentence to a subsequent sentence. The word *one* in sentence (2) illustrates cohesion based on substitution and the word *do* in sentence (4) illustrates cohesion based on ellipsis.

Substitution

(1) Did you ever find a lawnmower?
(2) Yes, I borrowed *one* from my neighbor.

Ellipsis

(3) Do you want to go with me to the store?
(4) Yes, I *do.*

The remaining three categories include the bulk of explicit cohesive ties in written English. The categories of *reference* and *conjunction* contain ties that are both grammatical and lexical. *Lexical reiteration and collocation* is restricted to ties which are presumably only lexical.

Reference cohesion occurs when one item in a text points to another element for its interpretation. Reference ties are of three types: *pronominals, demonstratives and definite articles,* and *comparatives.* Each of the sentence pairs below illustrates a different type of reference cohesion.

Reference Cohesion (Pronominal)

(5) At home, my father is himself.
(6) *He* relaxes and acts in *his* normal manner.

Reference Cohesion (Demonstratives)

(7) We question why they tell us to do things.
(8) *This* is part of growing up.

Reference Cohesion (Definite Article)

(9) Humans have many needs, both physical and intangible.
(10) It is easy to see *the* physical needs such as food and shelter.

Reference Cohesion (Comparatives)

(11) The older generation is often quick to condemn college students for being carefree and irresponsible.
(12) But those who remember their own youth do so *less* quickly.

The interpretation of the underlined elements in sentences (6), (8), (10), and (12) depends in each case upon presupposed information contained in the sentences immediately above it.

A fourth major class of cohesive ties frequent in writing is *conjunction.* Conjunctive elements are not in themselves cohesive, but they do "express certain meanings which presuppose the presence of other components in the discourse" (p. 226). Halliday and Hasan distinguish five types of conjunctive cohesion—*additive, adversative, causal, temporal,* and *continuative.* Examples of these subclasses of conjunctive cohesion appear below and illustrate how conjunctive cohesion extends the meaning of one sentence to a subsequent one.

Conjunctive Cohesion (Additive)

(13) No one wants to be rejected.
(14) *And* to prevent rejection we change our behavior often.

Conjunctive Cohesion (Adversative)

(15) Small children usually change their behavior because they want something they don't have.
(16) Carol, *however,* changed her behavior because she wanted to become part of a new group.

Conjunctive Cohesion (Causal)

(17) Today's society sets the standards.
(18) The people more or less follow it [sic].
(19) *Consequently,* there exists the right behavior for the specific situation at hand.

Conjunctive Cohesion (Temporal)

(20) A friend of mine went to an out-of-state college.
(21) *Before* she left, she expressed her feelings about playing roles to win new friends.

Conjunctive Cohesion (Continuative)

(22) Different social situations call for different behaviors.
(23) This is something we all learn as children and we, *of course,* also learn which behaviors are right for which situations.

Coordinating conjunctions (such as *and, but,* and *so*), conjunctive adverbs (such as *however, consequently,* and *moreover*), and certain temporal adverbs and subordinating conjunctions (such as *before, after,* and *now*) supply cohesive ties across sentence boundaries.

The last major class of cohesive ties includes those based on *lexical* relationships. *Lexical* cohesion differs from *reference* cohesion and *conjunctive* cohesion because every lexical item is potentially cohesive and because nothing in the occurrence of a given lexical item necessarily makes it cohesive. If we were to encounter the word *this* in a text, we would either supply a referent from our working memory of the text or reread the text to find a referent. Similarly, when we encounter a conjunctive adverb such as *however,* we attempt to establish an adversative relationship between two text elements. In contrast, lexical cohesion depends on some "patterned occurrence of lexical items" (p. 288). Consider the following sentences adapted from a mountaineering guidebook:

(24) The ascent up the Emmons Glacier on Mt. Rainier is long but relatively easy.
(25) The only usual problem in the climb is finding a route through the numerous crevasses above Steamboat Prow.
(26) In late season a *bergschrund* may develop at the 13,000-foot level, which is customarily bypassed to the right.

Three cohesive chains bind together this short text. The first chain (*ascent, climb, finding a route, bypassed to the right*) carries the topic—the way up the mountain. The second and third chains give the setting (*Glacier, crevasses, bergschrund*) (*Mt. Rainier, Steamboat Prow, 13,000-foot level*). These chains give clues to the interpretation of unfamiliar items. For most readers, *Steamboat*

Prow is unknown, but one can infer that it is a feature on Mt. Rainier. Similarly, *bergschrund* is a technical term referring to a crevasse at the head of a glacier where the moving ice breaks apart from the stationary ice clinging to the mountain. In this text, a reader can infer that *bergschrunds* are associated with glaciers and that they present some type of obstacle to climbers, even without the final clause in (26).

Lexical cohesion is the predominant means of connecting sentences in discourse. Halliday and Hasan identify two major subclasses of lexical cohesion: *reiteration* and *collocation*. *Reiteration* is in turn divided into four subclasses, ranging from repetition of the *same item* to repetition through the use of a *synonym or near-synonym*, a *superordinate item*, or a *general item*.

Lexical reiteration is usually easy to identify. An example of synonomy occurs in (25) and (26) with the pairing of *ascent* and *climb*. The three other subclasses are illustrated in the following student example:

Lexical Reiteration (Same Item), (Superordinate), and (General Item)

(27) Some professional tennis players, for example, grandstand, using obscene gestures and language to call attention to themselves.

(28) Other *professional athletes* do similar *things*, such as spiking a football in the end zone, to attract *attention*.

In (28), *professional athletes* is, in this case, a superordinate term for *professional tennis players*. Professional athletes in other sports are encompassed by the term. *Things*, in contrast, is a general term. Here *things* is used to refer anaphorically to two behaviors, "using obscene gestures and language." While superordinates are names of specific classes of objects, general terms are even more inclusive, not restricted to a specific set of objects. The other type of lexical reiteration, illustrated by sentences (27) and (28), is same-item repetition: *attention* is simply repeated.

All the lexical cohesive relationships which cannot be properly subsumed under lexical reiteration are included in a "miscellaneous" class called *collocation*. Collocation refers to lexical cohesion "that is achieved through the association of lexical items that regularly co-occur" (p. 284). Lexical cohesion through collocation is the most difficult type of cohesion to analyze because items said to collocate involve neither repetition, synonomy, superordination, nor mention of general items. What is important is that the items said to collocate "share the same lexical environment" (p. 286). The following student example illustrates this principle:

Lexical Cohesion (Collocation)

(29) On a camping trip with their parents, teenagers willingly do the household chores that they resist at home.

(30) They gather *wood for a fire*, help put up the *tent*, and carry *water from a creek or lake*.

Although the underlined items in (30) are presented as the "camping trip" equivalents of *household chores*, the cohesion between sentences (29) and (30)

results more directly from the associations of the underlined items with *camping trip.* The underlined items in sentence (30) collocate with *camping trip* in sentence (29). The mountaineering guidebook passage, however, is much more difficult to analyze. For one of the authors of the present article, antecedent knowledge of mountaineering allows *Steamboat Prow* to collocate with *Mt. Rainier* and *bergschrund* to collocate with *glacier.* For the other author, neither pair is lexically related by collocation apart from the text where they are connected by inference. We will return to this problem later in this essay.

In addition to the taxonomy that allows cohesive ties to be classified according to function, Halliday and Hasan introduce a second taxonomy. This second taxonomy allows cohesive ties to be classified according to the amount of text spanned by the *presupposed* and *presupposing* elements of a given tie. Halliday and Hasan posit four such "text-span" classes. Membership in a class is determined by the number of T-units a given cohesive tie spans.[7] Taken together, the two taxonomies Halliday and Hasan present allow any given cohesive tie to be classified in two different ways, one according to function and one according to distance. The four "text-span" classes contained in Halliday and Hasan's second taxonomy are illustrated in the following paragraph from a student paper:

Test-Span Classes (Immediate, Mediated, Remote, Mediated-Remote)

(31) *Respect* is one reason people change their behavior.
(32) For example, one does not speak with his *boss* as he would talk to a friend or co-worker.
(33) One might use four-letter words in talking to a co-worker, but probably not in talking to his *boss.*
(34) In talking to teachers or *doctors,* people also use bigger words than normal.
(35) Although the situation is different than when one speaks with a *boss* or a *doctor,* one often talks with a minister or priest different [sic] than he talks with friends or *family.*
(36) With the *family,* most people use a different language when they talk to parents or grandparents than when they talk to younger brothers and sisters.
(37) People's ability to use language in different ways allows them to show the *respect* they should toward different people, whether they are professionals, *family* members, clergy, friends and co-workers, or *bosses.*

Immediate cohesive ties semantically linked adjacent T-units. The repetition of *doctor* in sentences (34) and (35) creates an *immediate* tie, forcing the reader to assimilate the content of (34) into the content of (35). In contrast, the repetition of *family* in sentences (35), (36), and (37) forms a *mediated* tie. The semantic bridge established by the occurrence of *family* in (35) and (37) is channelled through or mediated by the repetition of *family* in (36). The cohesive tie involving the repetition of *family* is not simply a series of immediate ties, because once a lexical item appears in a text all subsequent uses

of that item presuppose the first appearance. *Immediate* and *mediated* ties join items in adjacent T-units. Such ties enable writers to introduce a concept in one T-unit and to extend, modify, or clarify that concept in subsequent and successive T-units.

Remote ties, on the other hand, result when the two elements of a tie are separated by one or more intervening T-units. The tie between *respect* in (31) and (37) is *remote;* here the repetition of the word signals to the reader that the semantic unit represented by the paragraph is now complete. Finally, ties which are both mediated and remote are called *mediated-remote.* An example of this type of cohesive tie appears in the repetition of *bosses* in sentences (32), (33), (35), and (37). Here the presupposing *bosses* in (37) is separated from the presupposed *boss* in (32) by intervening T-units (34) and (36) which contain no element relevant to the particular cohesive tie. Thus the tie is *remote.* However, the presupposing *bosses* is also *mediated* through repetitions of *boss* in (33) and (35). Hence the term *mediated-remote.* Skilled writers use mediated-remote ties to interweave key "themes" within the text.

Analysis of Student Essays

To explore the usefulness of Halliday and Hasan's theory of cohesion in writing research, we used their two taxonomies in an analysis of ten student essays. These essays were written by beginning University of Texas freshmen on the "changes in behavior" topic used in the Miami University sentence-combining experiment.[8] From 90 essays which had been rated holistically by two readers on a four-point scale, we selected five essays given the lowest score by both raters and five essays given the highest score. We analyzed these ten essays according to categories of error and according to syntactic features, as well as according to the number and types of cohesive ties. Our analyses of error and content variables yielded results similar to those other researchers have reported—that high-rated essays are longer and contain larger T-units and clauses, more nonrestrictive modifiers, and fewer errors.[9]

We anticipated that an analysis of cohesive ties in the high- and low-rated essays would reveal similar gross differences. The results of our analysis confirmed this expectation. At the most general level of analysis, the high rated essays are much more dense in cohesion than the low-rated essays. In the low-rated essays, a cohesive tie of some type occurs once every 4.9 words; in the high-rated essays, a tie occurs once every 3.2 words, a difference in mean frequency of 1.7 words. Likewise, a large difference in the mean number of cohesive ties per T-unit appears, with 2.4 ties per T-unit in the low-rated essays and 5.2 ties per T-unit in the high-rated essays. The figures for this and the preceding index, however, are not precisely comparable because the T-units in the high-rated essays are, on the average, 1.64 words longer than those in the low-rated essays. By dividing the number of cohesive ties in an essay set by the number of words in that set, we arrived at another general

index of cohesive density. In the high-rated essays, 31.7% of all words contribute to explicit cohesive ties while only 20.4% of the words in the low-rated essays contribute to such ties.

The ways in which writers of the high- and low-rated essays form cohesive ties also distinguish the two groups of five essays from each other. Writers of the high-rated essays use a substantially higher relative percentage of *immediate* (High: 41.6%/Low: 32.8%) and *mediated* (High: 7.6%/Low: 0.8%) cohesive ties than do the writers of the low-rated essays. On the other hand, writers of the low-rated essays use more *mediated-remote* (High: 25.9%/Low: 36.7%) and *remote* ties (High: 26.9%/Low: 29.7%). These percentages allow us to focus on some crucial differences between the two essay sets. The larger relative percentage of *immediate* cohesive ties in the high-rated essays suggests, among other things, that the better writers tend to establish stronger cohesive bonds between individual T-units than do the writers of the low-rated essays. Analyses of *reference* and *conjunctive cohesion* support this observation. Writers of high-rated essays employ reference cohesion about twice as often, 84.1 times to 47.8 times per 100 T-units, as the writers of low-rated papers. The largest difference in the occurrence of referential cohesion is reflected in the higher frequency of third-person pronouns in the high-rated essays (High: 25.1 per 100 T-units/Low: 5.1 per 100 T-units). This lower frequency of third-person pronouns in the low-rated essays may be a direct result of the less skilled writers' attempts to avoid errors such as ambiguous pronoun reference. Because third-person pronouns usually refer back to the T-unit immediately preceding, we can infer that the writers of high-rated essays more often elaborate, in subsequent and adjacent T-units, topics introduced in a given T-unit.

Also contributing importantly to the greater use of *immediate* cohesive ties is the frequency with which the more skillful writers use *conjunction* to link individual T-units. Conjunctive ties most often result in *immediate* cohesive ties between T-units. It is not surprising, then, to find that the writers of high-rated essays employ over three times as many conjunctive ties (High: 65.4 per 100 T-units/Low: 20.4 per 100 T-units) as the writers of low-rated essays. Neither is it surprising to discover that the more skillful writers employ all five types of conjunction while the less skillful writers use only three. As is the case with pronominal references that cross T-unit boundaries, conjunctives are most often used to extend concepts introduced in one T-unit to other T-units which follow immediately in the text. Thus the more skillful writers appear to extend the concept introduced in a given T-unit considerably more often than do the less skillful writers. One major effect of such semantic extensions is, of course, essay length; and this finding helps to explain why the high-rated essays are, on the average, 375 words longer than the low-rated essays.

The relative frequency of *lexical cohesion* gives another indication that the writers of high-rated essays are better able to expand and connect their ideas

than the writers of the low-rated essays. By far the largest number of cohesive ties, about two-thirds of the total ties for both the high and low samples, fall into the general category of *lexical cohesion*. Writers of the high-rated essays create some type of lexical tie 340 times per 100 T-units or every 4.8 words. Writers of the low-rated essays, however, manage a lexical tie just 161 times per 100 T-units or every 7.4 words. The majority of lexical ties (65%) in the low essays are repetitions of the same item. This distribution is reflected to a smaller degree in the high essays, where 52% of the total lexical ties fall into the *same item* subcategory. Writers of high-rated essays, however, form many more lexical collocations. Lexical collocations appear 94 times per 100 T-units in the high-rated essays in contrast to 28.8 times per 100 T-units in the low-rated essays.

Cohesion and Invention

These cohesion profiles suggest to us an important difference between the invention skills of the two groups of writers. The better writers seem to have a better command of invention skills that allow them to elaborate and extend the concepts they introduce. The poorer writers, in contrast, appear deficient in these skills. Their essays display a much higher degree of lexical and conceptual redundancy. The high percentage of lexical redundancy and the low frequency of lexical collocation in the low-rated essays are indications of this difference. The text-span categories also point to this difference. In the low-rated essays two-thirds of the cohesive ties are interrupted ties—*mediated-remote* or *remote* ties—which reach back across one or more T-units, indicating that the writers of the low-rated essays generally fail to elaborate and extend concepts through successive T-units.

The larger proportion of interrupted ties in the low-rated papers strongly suggests that substantially less new information or semantic content is introduced during the course of a low-rated essay than during the course of a high-rated essay. If more new information had been introduced in the low-rated essays, the writers would have had to rely more heavily than they did on *immediate* and *mediated* cohesive ties in order to integrate, to weave, the new information into the text. The writers of the low-rated papers tend more toward reiteration of previously introduced information than do the writers of the high-rated papers. Indeed, in reading the low-rated essays one can not help noting a good deal of what might be called conceptual and lexical redundancy. The following example illustrates this characteristic:

> Some people have to change their behavior around different acquaintances. One reason is that they want to make a good impression on others. You have to act different in front of a person who is giving you a job interview because you want to make a good impression. You, most of the time, act differently to fit in a crowd. You will change your behavior to

> get people to like you. You change your behavior to agree with peoples [sic] in the crowd.

This paragraph from a low-rated paper has a fairly strong beginning: it states a topic in the first sentence, modifies that topic in the second sentence, illustrates the topic in the third sentence, and gives another example in the fourth sentence. The next two sentences, however, simply reiterate what is said in the fourth sentence. The principal lexical items in the last two sentences—*change, behavior, people,* and *crowd*—are repetitions of items introduced earlier in the paragraph and offer little new information. Although for purposes of attaining cohesion in a text some redundancy is a virtue, the redundancy in the low-rated essays seems to be a flaw because these writers failed to supply additional information at the point where it would be expected to appear. Had this additional information been supplied, the writers would have had to use *immediate* and *mediated* ties in order to connect it to the rest of the text.

Compare the previous example paragraph from a low-rated paper with the following paragraph from a high-rated paper.

> It is a job that really changes our behavior. Among other changes, we change the way we dress. In many jobs college graduates want to look responsible and mature, projecting an image of competence. The college student who wore faded blue jeans is now in three-piece suits. He feels the need to be approved of and accepted by his boss and associates. While he talked of socialism in college, he now reaps the profits of capitalism. While in college he demanded honesty in the words and actions of others, on the job he is willing to "kiss ass" to make friends or get a promotion. Indeed, working can change behavior.

Notice that in the paragraph from the high-rated paper, *behavior* is repeated only one time. Yet the reader never questions that the paragraph is about changes in behavior. The writer repeatedly supplies examples of types of behavior, which are linked to the topic by a series of lexical collocations (e.g., *behavior, dress, look responsible, blue jeans, three-piece suits*). Clearly, the paragraph from the high-rated paper extends the semantic domain of the concept *behavior* to include a number of differentiated lexical items. Low-rated papers rarely show such extended series of collocations.

Analyses of cohesion thus measure some aspects of invention skills. The low-rated essays stall frequently, repeating ideas instead of elaborating them. Our analyses also suggest that the writers of the low-rated papers do not have working vocabularies capable of extending, in ways prerequisite for good writing, the concepts and ideas they introduce in their essays. Indeed, skill in invention, in discovering what to say about a particular topic, may depend in ways yet unexplored on the prior development of adequate working vocabularies. If students do not have in their working vocabularies the lexical items required to extend, explore, or elaborate the concepts they introduce, practice in invention can have only a limited effect on overall writing quality.

Our analyses further point to the underdevelopment of certain cognitive skills among the writers of the low-rated papers. The low-rated papers not

only exhibit a great deal of redundancy, but (as noted earlier) also include relatively fewer *conjunctive* and *reference* ties and *immediate* and *mediated* ties. Besides lacking adequate vocabularies, writers of the low-rated essays seem to lack in part the ability to perceive and articulate abstract concepts with reference to particular instances, to perceive relationships among ideas, and to reach beyond the worlds of their immediate experience.

All this is to suggest that analyses of cohesion may be potentially useful in distinguishing between stages of writing development. Clearly, cohesion analyses measure more sophisticated aspects of language development than do error analyses and syntactic analyses. Cohesion analyses also give us some concrete ways of addressing some of the differences between good and poor writing, differences which heretofore could not be explained either to ourselves or to our students in any but the most abstract ways. We thus anticipate that Halliday and Hasan's taxonomies can be usefully applied in developmental studies as well as in studies such as the present one.

Cohesion, Coherence, and Writing Quality

However promising cohesion analysis appears as a research tool and however encouraging the results of the present study seem, we feel that a number of important questions cannot be answered by analyzing cohesion. The first of these questions concerns writing quality. The quality or "success" of a text, we would argue, depends a great deal on factors outside the text itself, factors which lie beyond the scope of cohesion analyses. Recall that Halliday and Hasan exclude *exophoric,* or outside-text, references from their taxonomy of explicit cohesive ties. We think that writing quality is in part defined as the "fit" of a particular text to its context, which includes such factors as the writer's purpose, the discourse medium, and the audience's knowledge of an interest in the subject—the factors which are the cornerstones of discourse theory and, *mutatis mutandis,* should be the cornerstones of research in written composition.[10] We are not alone in this view. Several students of written discourse—among them Joseph Grimes,[11] Teun van Dijk, [12] Nils Enkvist,[13] and Robert de Beaugrande[14]—distinguish *cohesion* and *coherence.* They limit cohesion to explicit mechanisms in the text, both the types of cohesive ties that Halliday and Hasan describe and other elements that bind texts such as parallelism, consistency of verb tense, and what literary scholars have called "point of view."[15] Coherence conditions, on the other hand, allow a text to be understood in a real-world setting. Halliday and Hasan's theory does not accommodate real-world settings for written discourse or, consequently, the conditions through which texts become coherent. We agree with Charles Fillmore's contention that

> the scenes . . . [audiences] construct for texts are partly justified by the lexical and grammatical materials in the text and partly by the interpreter's own contributions, the latter being based on what he knows about

> the current context, what he knows about the world in general, and what he assumes the speaker's intentions might be.[16]

Hence lexical collocations within a text are understood through cues which the writer provides and through the reader's knowledge of general discourse characteristics and of the world to which the discourse refers.

Thus lexical collocation is in all likelihood the subcategory of cohesion that best indicates overall writing ability, as well as disclosing distinctions among written texts that represent different discourse modes and purposes. An examination of lexical cohesive ties shows how writers build ideas, how they are able to take advantage of associations to weave together a text. But a fundamental problem lies in the analysis of a writer's text. Whose collocations do we analyze—the reader's or the writer's? One simple proof that the two do not always coincide can be found in the unintentional sexual references that students occasionally produce—the kind that get passed around the faculty coffee room.

Consider again the mountaineering guidebook passage in sentences (24), (25), and (26). We have already established that for mountaineers and glaciologists, *bergschrund* probably collocates with *glacier,* but for many other persons the two items do not collocate. Yet a naive reader presented this text probably would not stop to consult a dictionary for the lexical item, *bergschrund,* but would infer from its context that it is some type of obstacle to climbers and continue reading. Herbert Clark theorizes that we comprehend unknown items like *bergschrund* by drawing inferences.[17] We make inferences on the basis of what we can gather from the explicit content and the circumstances surrounding a text, through a tacit contract between the writer and reader that the writer will provide only information relevant to the current topic. In the case of the mountaineering passage, the circumstances of the text greatly affect our understanding of it. The type of text—a guidebook—follows a predictable organization, what has been called a *script* in research on artificial intelligence.[18] The guidebook contains a series of topics with a clear, yet implicit, goal: to inform the reader how to get to the top of a mountain. We expect the author to give us only information relevant to the particular route. Accordingly, readers understand *bergschrund* as an obstacle through a combination of cues—overt signals in the text such as the parallelism of the *bergschrund* sentence with the sentence about crevasses above it and, for those readers familiar with the type of text, implicit signals such as the following of the guidebook "script." Although Halliday and Hasan do not include parallelism in their taxonomy, parallelism often creates a cohesive tie.

Cohesion and coherence interact to a great degree, but a cohesive text may be only minimally coherent. Thus cohesion-based distinctions between texts rated high and low in quality can be misleading. Besides explicit links within a text, a text must conform to a reader's expectations for particular types of texts and the reader's knowledge of the world. A simple example will illustrate this point:

(38) The quarterback threw the ball toward the tight end.
(39) Balls are used in many sports.
(40) Most balls are spheres, but a football is an ellipsoid.
(41) The tight end leaped to catch the ball.

Sentences (39) and (40), while cohesive, violate a coherence condition that the writer provide only information relevant to the topic. The major problem with this short text is that a reader cannot construct what Fillmore calls a real-world scene for it; that is, the text neither seems to have a clear purpose nor appears to meet the needs of any given audience. Because it has no clear purpose, it lacks coherence, in spite of the cohesive ties which bind it together. In addition to a cohesive unity, written texts must have a pragmatic unity, a unity of a text and the world of the reader. A description of the fit of a text to its context, as well as descriptions of what composition teachers call writing quality, must specify a variety of coherence conditions, many of them outside the text itself.

Implications for the Teaching of Composition

One implication of the present study is that if cohesion is better understood, it can be better taught. At present, in most college writing classes, cohesion is taught, explicitly or implicitly, either through exercises, classroom instruction, or comments on student papers. Many exercises not explicitly designed to teach cohesion do in fact demand that students form cohesive ties. Open sentence-combining exercises, for example, offer as much practice in forming cohesive ties as they do in manipulating syntactic structures, a fact which may explain the success of certain sentence-combining experiments as well as the failure of research to link syntactic measures such as T-unit and clause length to writing quality.[19] An open sentence-combining exercise about Charlie Chaplin might contain a series of sentences beginning with the name *Charlie Chaplin.* Such an exercise would, at the very least, demand that students change most of the occurrences of *Charlie Chaplin* to *he* in order to produce an acceptable text. Students working either from contextual cues or from their knowledge of Chaplin might also use phrases like *the comic genius* or *the little tramp* to substitute for the proper name *Chaplin.*

If cohesion is often implicitly incorporated in writing curricula, coherence is often ignored. A great portion of the advice in composition textbooks stops at sentence boundaries. Numerous exercises teach clause and sentence structure in isolation, ignoring the textual, and the situational, considerations for using that structure. The passive is a classic example:

(42) The police apprehended the suspect as he left the bank.
(43) He is being held in the county jail.
(43a) The police are holding the suspect in the county jail.

A student following her teacher's advice to avoid the passive construction

might revise sentence (43) to (43a). If she did so, she would violate the usual sequence of information in English, where the topic or "old" information is presented first.[20] In active sentences, such as (43a), where the object expresses the topic, a revision to the passive is often preferable. Avoiding the passive with (43a) would also require the unnecessary and uneconomical repetition of *police* and *suspect.* Consequently, maxims such as "Avoid passives" ignore the coherence conditions that govern the information structure of a text.

Other discourse considerations are similarly ignored in traditional advice on how to achieve coherence. As E. K. Lybert and D. W. Cummings have observed, the handbook injunction "Repeat key words and phrases" often *reduces* coherence.[21] Our analysis of cohesive ties in high- and low-rated essays substantiates Lybert and Cummings' point. While the low-rated papers we examined contain fewer cohesive ties than the high-rated papers in equivalent spans of text, the low-rated papers rely more heavily on lexical repetition. Also contrary to a popular notion, frequent repetition of lexical items does not necessarily increase readability. Roger Shuy and Donald Larkin's recent study shows lexical redundancy to be a principal reason why insurance policy language is difficult to read.[22]

Our analysis of cohesion suggests that cohesion is an important property of writing quality. To some extent the types and frequencies of cohesive ties seem to reflect the invention skills of student writers and to influence the stylistic and organizational properties of the texts they write. However, our analysis also suggests that while cohesive relationships may ultimately affect writing quality in some ways, there is no evidence to suggest that a large number (or a small number) of cohesive ties of a particular type will positively affect writing quality. All discourse is context bound—to the demands of the subject matter, occasion, medium, and audience of the text. Cohesion defines those mechanisms that hold a text together, while coherence defines those underlying semantic relations that allow a text to be understood and used. Consequently, coherence conditions—conditions governed by the writer's purpose, the audience's knowledge and expectations, and the information to be conveyed—militate against prescriptive approaches to the teaching of writing. Indeed, our exploration of what cohesion analyses can and cannot measure in student writing points to the necessity of placing writing exercises in the context of complete written texts. Just as exclusive focus on syntax and other formal surface features in writing instruction probably will not better the overall quality of college students' writing, neither will a narrow emphasis on cohesion probably produce significantly improved writing.[23]

Notes

1. Most notably, Mina Shaughnessy, *Errors and Expectations,* (New York: Oxford University Press, 1977).

2. See Ann O. Gebhard, "Writing Quality and Syntax: A Transformational Analysis of Three Prose Samples," *Research in the Teaching of English,* 12 (October, 1978), 211-231.

3. Composition theorists, however, have not stopped at sentence boundaries. Several of the early efforts to describe relationships across sentences are summarized in Richard L. Larson, "Structure and Form in Non-Fiction Prose," in *Teaching Composition: Ten Bibliographical Essays,* ed. Gary Tate (Fort Worth, Texas: Texas Christian University Press, 1976), pp. 45-71. The reader will note certain similarities between Halliday and Hasan's taxonomy of cohesive ties set out later in the present essay and the work of previous composition theorists.

4. "Measuring Syntactic Growth: Errors and Expectations in Sentence-Combining Practice with College Freshmen," *Research in the Teaching of English,* 12 (October, 1978), 233-244.

5. No comprehensive overview of work in discourse in all of these fields exists at present. Extensive bibliographies, however, can be found in *Current Trends in Textlinguistics,* ed. Wolfgang U. Dressler (Berlin: de Gruyter, 1978) and Robert de Beaugrande, *Text, Discourse, and Process* (Norwood, N.J.: Ablex, 1980).

6. (London: Longman, 1976).

7. The term *T-unit,* of course, comes from Kellogg Hunt's *Grammatical Structures Written at Three Grade Levels,* NCTE Research Report No. 3 (Champaign, Ill.: National Council of Teachers of English, 1965). Hunt defined a *T-unit* as an independent clause and all subordinate elements attached to it, whether clausal or phrasal. Halliday and Hasan do not use the term *T-unit,* but they do define their four "text-span" classes according to the number of simple and complex sentences that the presupposing element of a cohesive tie must reach across for the presupposed element (see pp. 340-355). There is good reason to define the four "text-span" classes in terms of T-units. To examine only cohesive ties that span the boundaries of orthographic sentences would ignore the large number of conjunctive relationships, such as addition and causality, between independent clauses.

8. Max Morenberg, Donald Daiker, and Andrew Kerek, "Sentence Combining at the College Level: An Experimental Study," *Research in the Teaching of English,* 12 (October, 1978), 245-256. This topic asks students to write about why we act differently in different situations, using specific illustrations from personal experience.

9. Detailed analyses for data summarized in this section are reported in Stephen P. Witte and Lester Faigley, *A Comparison of Analytic and Synthetic Approaches to the Teaching of College Writing,* Unpublished manuscript. The high-rated essays are, on the average, more than twice as long as the low-rated essays (647 words/270 words). Errors in three major categories were counted—punctuation, spelling, and grammar. Grammatical errors include errors in verb tenses, subject-verb agreement, pronoun reference, pronoun number agreement, and dangling or misplaced modifiers. The low-rated essays exhibit an error of some type nearly three times as often as the high-rated essays—one every 29 words as opposed to one every 87 words. For example, errors in end-stop punctuation, resulting in either a comma splice or a fragment, occur nearly eight times as often in the low-rated essays as in the high-rated essays. Misspelled words are over four times as frequent in the low-rated essays, and grammatical errors appear over twice as often.

Syntactic comparisons were made according to the number of words per T-unit and clause, and according to the frequency and placement of nonrestrictive or "free" modifiers. The low-rated essays contain T-units and clauses considerably shorter than the high-rated essays (High: 15.3 words per T-unit, 9.3 words per clause/Low: 13.7 words per T-unit, 7.5 words per clause). Nonrestrictive modifiers in all positions—initial, medial, and final—appear in the T-units of the high-rated essays nearly three times as frequently as in the low-rated essays (High: 28.5% of all T-units contain nonrestrictive modifiers/Low: 10.1% contain nonrestrictive modifiers). The high-rated essays also have twice the percentage of total words in nonrestrictive modifiers that the low-rated essays have.

10. Stephen P. Witte, "Toward a Model for Research in Written Composition," *Research in the Teaching of English,* 14 (February, 1980), 73-81.

11. *The Thread of Discourse* (The Hague: Mouton, 1975).

12. *Text and Context* (London: Longman, 1977).

13. "Coherence, Pseudo-Coherence, and Non-Coherence," in *Reports on Text Linguistics: Semantics and Cohesion,* ed. Jan-Ola Ostman (Abo, Finland: Research Institute of the Abo Akademi Foundation, 1978), pp. 109-128.

14. "The Pragmatics of Discourse Planning," *Journal of Pragmatics,* 4 (February, 1980), 15-42.

15. Susumo Kuno describes some linguistic features controlled by point of view, which he calls "empathy." See "Subject, Theme and the Speaker's Empathy—A Reexamination of Relativization Phenomena," in *Subject and Topic,* ed. Charles N. Li (New York: Academic Press, 1976), pp. 417-444.

16. "Topics in Lexical Semantics," in *Current Issues in Linguistic Theory,* ed. Roger W. Cole (Bloomington: Indiana University Press, 1977), p. 92.

17. "Inferences in Comprehension," in *Basic Processes in Reading: Perception and Comprehension,* ed. David LaBerge and S. Jay Samuels (Hillsdale, N.J.: Erlbaum, 1977), pp. 243-263.

18. R. C. Schank and R. P. Abelson, *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures* (Hillsdale, N.J.: Erlbaum, 1977).

19. See Lester Faigley, "Names in Search of a Concept: Maturity, Fluency, Complexity, and Growth in Written Syntax," *College Composition and Communication,* 31 (October, 1980), 291-300.

20. The theme-rheme distinction is the product of the Prague school of linguistics. Other researchers use the terms "topic-comment" or "given-new" information to refer to essentially the same concept. See Vilem Mathesius, *A Functional Analysis of Present Day English on a General Linguistic Basis,* ed. Josef Vachek (Prague: Academia, 1975). Also relevant are Wallace L. Chafe, *Meaning and the Structure of Language* (Chicago: University of Chicago Press, 1970); M. A. K. Halliday, "Notes on Transitivity and Theme in English: II," *Journal of Linguistics,* 3 (October, 1967), 199-244; Herbert Clark and Susan Haviland, "Comprehension and the Given-New Contract," in *Discourse Production and Comprehension,* ed. Roy Freedle (Norwood, N.J.: Ablex, 1977), pp. 1-40; and Liisa Lautamatti, "Observations on the Development of Topic in Simplified Discourse," in *Text Linguistics, Cognitive Learning, and Language Teaching,* ed. Viljo Kohonen and Nils Erik Enkvist (Turku, Finland: University of Turku, 1978), pp. 71-104.

21. "On Repetition and Coherence," *College Composition and Communication,* 20 (February, 1969), 35-38.

22. "Linguistic Consideration in the Simplification/Clarification of Insurance Policy Language," *Discourse Processes,* 1 (October-December, 1978), 305-321.

23. John Mellon and James Kinneavy both make the point that while the Miami University sentence-combining students improved significantly in writing quality, they may not have done so because they learned to manipulate syntactic structures better, but because they were taught to put together complete texts. See "Issues in the Theory and Practice of Sentence Combining: A Twenty-Year Perspective," in *Sentence Combining and the Teaching of Writing,* ed. Donald Daiker, Andrew Kerek, and Max Morenberg (Akron, Oh.: University of Akron, 1979), p. 10; and "Sentence Combining in a Comprehensive Language Framework" in the same volume, p. 66.

WPA: Writing Program Administration

Beginning with its fall, 1979 issue, the *WPA Newsletter* became a refereed journal, *WPA: Writing Program Administration.* It is published three times a year by the Council of Writing Program Administrators and by the School of Humanities and Office of Publications, Brooklyn College—City University of New York. Contributions should be addressed to Kenneth A. Bruffee, Editor *WPA,* English Department, Brooklyn College, Brooklyn, NY 11210. Correspondence concerning subscriptions (included in membership in the Council of Writing Program Administrators) should be addressed to Joseph Comprone, WPA Treasurer, English Department, University of Louisville, Louisville, KY 40208.

http://www.jstor.org

LINKED CITATIONS

You have printed the following article:

Coherence, Cohesion, and Writing Quality

Stephen P. Witte; Lester Faigley

College Composition and Communication, Vol. 32, No. 2, Language Studies and Composing. (May, 1981), pp. 189-204.

Stable URL:

http://links.jstor.org/sici?sici=0010-096X%28198105%2932%3A2%3C189%3ACCAWQ%3E2.0.CO%3B2-E

This article references the following linked citations. If you are trying to access articles from an off-campus location, you may be required to first logon via your library web site to access JSTOR. Please visit your library's website or contact a librarian to learn about options for remote access to JSTOR.

Notes

[19] **Names in Search of a Concept: Maturity, Fluency, Complexity, and Growth in Written Syntax**

Lester Faigley

College Composition and Communication, Vol. 31, No. 3. (Oct., 1980), pp. 291-300.

Stable URL:

http://links.jstor.org/sici?sici=0010-096X%28198010%2931%3A3%3C291%3ANISOAC%3E2.0.CO%3B2-Z

NOTE: *The reference numbering from the original has been maintained in this citation list.*