

Measuring the coherence of writing using topic-based analysis.

Watson Todd, R., Thienpermpool, P. and Keyuravong, S. Assessing Writing vol. 9 no. 2 pp. 85-104. (2004)

The definitive version of this article was published as Watson Todd, R., Thienpermpool, P. and Keyuravong, S. (2004)

Measuring the coherence of writing using topic-based analysis. Assessing Writing vol. 9 no. 2 pp. 85-104.

doi:10.1016/j.asw.2004.06.002

It is available at

<http://www.sciencedirect.com/science/article/pii/S1075293504000248>

Measuring the coherence of writing using topic-based analysis

Abstract

Among the many possible aspects to assess in writing, one of the most problematic is coherence. The problems with marking coherence arise because it is by nature subjective. However, the reasonable probability of several readers reaching a consensus concerning the coherence of a text suggests that it may be possible to assign relatively reliable marks for coherence. In this paper, topic-based analysis (Watson Todd, 1998) was chosen as a method of evaluating coherence and the results of this evaluation were compared with teachers' marks for coherence. Topic-based analysis involves identifying key concepts in a text, identifying the relationships between these concepts, linking the relationships into a hierarchy, and mapping the text onto the hierarchy. Doing this allows several different measures of coherence to be generated. Comparing these measures against the teachers' scores, it was found that the number of moves between key concepts per ten T-units correlated most closely with the teachers' marks. The results have implications concerning the basis of teachers' marks for coherence and the applicability of topic-based analysis as a method of assessing coherence of writing.

Keywords

Coherence

Writing assessment

Topic-based analysis

Marking

Measuring the coherence of writing using topic-based analysis

Writing may be assessed for a variety of purposes. Perhaps the most familiar to teachers is to evaluate certain aspects of learners' language ability, but writing assessment may also be conducted to evaluate non-linguistic abilities of learners, to evaluate courses and materials, and as an instrument in research. Whatever the purpose, a key issue to be considered is how the writing is to be scored. For most writing assessment, there are two main scoring options: holistic scoring and analytic scoring (Cohen, 1994; Weigle, 2002; Weir, 1990, 1993). The latter uses separate scales to measure different aspects of writing, such as content, cohesion, coherence, mechanics and so on. While it is possible to score some aspects (such as mechanics) objectively, others by their nature require subjective interpretations on the part of raters potentially leading to problems of reliability. In this paper, we will focus on one of these subjective aspects, namely coherence, and investigate whether a research method originally designed for the analysis of classroom discourse can be used as the basis of assessing coherence in students' writing. The research method chosen is relatively objective, and we will compare scores derived from applying it to students' writing with teachers' scores for coherence to examine its appropriateness as a basis for assessing coherence.

The assessment of coherence in writing

Most analytic scoring schemes for writing include descriptors concerning discourse-level features of writing such as coherence. For example, high-quality writing should be "well organized", exhibit "logical sequencing" and be "cohesive" (Jacobs et al., 1981 quoted in Weigle 2002: 116), should have "a clear progression of ideas well-linked" (Anderson, n. d. quoted in Hughes, 2003: 102), and should provide "clear and consistent evidence of the ability, (*sic*) to produce organised coherent and cohesive discourse" (UCLES/RSA Certificate in Communicative Skills quoted in Weir, 1990: 172). On the other hand, poor-quality writing may be "so fragmentary that comprehension of the intended communication is virtually impossible" (descriptor for Cohesion, TEEP Attribute Writing Scales quoted in Weir, 1990: 69). Such analytic descriptors aim to increase the reliability of scoring, even where no rater training is provided (Weir, 1990). While the purpose of using such descriptors is laudable, their effects on reliability are less clear for two reasons. Firstly, the features described, such as whether a piece of writing is well-organised or not, may require subjective interpretations on the part of the markers. Secondly, the descriptions themselves may lead to confusion. For

example, from a linguistic perspective, the description of poor quality cohesion from the TEEP scale given above would actually seem to be describing coherence rather than cohesion. Such confusions and lack of clarity may cause reliability problems, so let us look more closely at what exactly is meant by coherence, the feature of writing which is the focus of this paper.

Coherence is often contrasted with cohesion, where cohesion is "the mutual connection of components of surface text" (Bell, 1991: 165), whereas coherence refers to the less tangible ways of connecting discourse which are not overt and which reside in how people interpret texts rather than in the texts themselves (Yule, 1996). With cohesion, then, it is possible to point to certain elements in a text as providing cohesion and thus cohesion is a relatively objective quality of discourse. Coherence, on the other hand, is "subjective and judgments concerning it may vary from reader to reader" (Hoey, 1991: 12). In writing assessment, this subjectivity means that scores assigned for coherence are likely to be unreliable.

It would therefore seem that we are faced with a situation where we may wish to assign marks for coherence since it is a key feature of discourse, but where any marks assigned will be derived from subjective interpretations. However, even though coherence is subjective, Hoey (1991: 266) also states that an "overwhelming consensus" of opinion concerning the level of coherence of naturally-occurring discourse can be achieved. Such a statement can give us hope that some reasonably reliable method of scoring coherence is attainable, despite its inherent subjectivity. To find such a method, let us turn to the literature on discourse analysis and examine those methods frequently used to analyse coherence in discourse.

Analysing coherence

A wide range of methods of analysing coherence exists. To decide which one to use as the basis for assessing coherence in students' writing, we need to set up some criteria that can be used to select a method that suits our purpose of finding a relatively objective method of measuring coherence against which teachers' scores for coherence can be compared.

The first criterion for selecting a suitable method of analysis is that it should be relatively objective. This criterion rules out several methods of analysing coherence, especially those situated within a conversation analysis framework. For example, the work of Crow (especially 1983) aims to identify kinds of topic progression in discourse, and consequently

can be used to analyse coherence. In this approach, the researcher examines the discourse for points at which the topic drifts or shifts. Unfortunately for our purposes, the identification of such points relies on the researcher's subjective opinions and thus the method is unlikely to lead to a reliable measure of coherence.

The second criterion is that the method should unequivocally measure coherence rather than some other construct. Unfortunately this rules out topical structure analysis (Connor and Farmer, 1990; Lautamatti, 1978; Schneider and Connor, 1990) or theme-rheme progression (Daneš, 1974) which has previously been used in research on second language writing coherence (e.g. Lee, 2002). In this approach, the themes and rhemes of each sentence are identified and how succeeding themes and rhemes relate to each other provides an indication of the coherence of the discourse. Although this approach is temptingly objective, it is unclear whether the method is actually measuring coherence and so has construct validity. For example, in a comparison of six different methods of analysing coherence including topical structure analysis, the methods were compared to a control analysis that attempted to identify topics based on random principles (Watson Todd, 2003). It was found that a topical structure analysis produced results uncomfortably close to those produced by this randomly designed control analysis (whereas the other methods did not produce such close results) casting doubt on whether topical structure analysis really does provide a measure of coherence.

The third criterion is that the method should focus on the type of coherence most commonly found in writing. There are two main kinds of coherence: interactional coherence and propositional coherence (Stubbs, 1983). Interactional coherence occurs when succeeding speech acts in discourse are implicitly linked, and such coherence is predominant in informal spoken language. Propositional coherence, on the other hand, concerns the implicit links created by the ideational content of the discourse and is predominant in more formal settings and in written language (Lautamatti, 1990; Redeker, 1990). The method chosen should therefore focus on propositional coherence and methods, such as rhetorical structure analysis (Mann and Thompson, 1988), which focus on the functions within the discourse and thus concern interactional coherence, should not be used.

One method which does meet all three criteria is topic-based analysis (Watson Todd, 1997, 1998). In this approach, key concepts are identified primarily through frequency, the logical relationships between these concepts are identified, and from these relationships line diagrams

representing schemata of the discourse are drawn up. The coherence of the discourse can then be measured by mapping the discourse onto these line diagrams. Although parts of this approach, such as the identification of relationships between concepts, are somewhat subjective, it is far more objective than an approach based in conversation analysis. Watson Todd (2003) also found that it is likely to be a more valid measure of coherence than topical structure analysis, and by focusing on key concepts topic-based analysis is concerned with propositional, rather than interactional, coherence. In this paper, then, we will examine the reliability of topic-based analysis as a method of assessing coherence in writing.

Applying topic-based analysis to writing

The original work on topic-based analysis used spoken classroom discourse as the data for analysis. The method, however, would appear applicable to written discourse as well. In applying topic-based analysis to written language, we will generally follow the original methodology, but, where possible, we will try to make the analytical methods more objective.

To illustrate how the methodology of topic-based analysis can be applied to writing, we will use the following student essay as an example.

Text 1

Pollution in Thailand¹! /

Pollution² is usually defined as something that people³ produce in large enough quantities that it^(pollution) interferes with our health or well-being. / Two primary factors that affect the amount of damage done by pollution² are the size of population and the development of technology⁴ that invents new forms of pollution. /

Throughout history, humans³ have made numerous attempts to eliminate the misery caused by hunger and disease. / In general, we^(humans)³ rely on science and technology⁴ to improve quality of life. / However, technological progress often offers short-term solutions that in the process of solving one problem can create new forms of pollution². /

Nowadays, there are so many pollution's problems in Thailand¹. / Although we^(Thai people)⁶ have multiple battlefields / but the situation of Thailand's pollution¹ not yet better. / If you can remember the event that occur few year ago "Chaowphaya River" was the one of pollution's¹ victim. / I think Thai people⁶ don't collaborate enough to protect environment⁷. / People⁶ still throw away garbages⁸ / and the industrials⁹ do

that^(throw away garbage)8 too! / Thai's Law and organization that concern are not earnest enough to stop and punish them^(industrials)9. /

I think in the future Thailand⁵ may be the country of garbage⁸. / That time to restore environment in Thailand⁷ and get rid to pollution². /

The conventions used in presenting Text 1 are as follows:

/ indicates a break between T-units

Superscript numbers indicate key concepts

Superscript bracketed text indicates the likely referents of any referring expression

To conduct a topic-based analysis of Text 1, we need to apply the following six stages to the text.

Stage 1 Preparing the text for analysis

There are three ways in which a text needs to be prepared before we can conduct a topic-based analysis. First, the text needs to be broken into units. Since in this study we are investigating the writing of English learners whose use of sentence structure and punctuation may be inappropriate, the unit chosen is the T-unit. Originally defined by Hunt (1970), a T-unit is an independent clause together with all related dependent clauses (Fries, 1994). It is relatively straightforward to identify independent clauses and to assign dependent clauses to independent clauses in students' writing.

Second, ellipted material is identified. In this study, ellipted material is taken as being material which the writer appears to perceive as being readily available to readers but which is not explicitly stated in the T-unit under consideration. Generally, material is taken as being ellipted where T-units are syntactically incomplete, and parallel structure is the prime basis for identifying what material has been ellipted. For example, the following sentence ends with a transitive verb with no object and therefore it is assumed that the object is ellipted and may be identified from parallel structure (the ellipted material is indicated as subscript bracketed text): *When they want to use water, they use* _(water).

Third, the referents of any referring expressions are identified. In doing this, preference is given to a referent which matches to the same syntactically and semantically relevant position in another nearby T-unit (Sotillo, 1999). For example, in the following extract, *it* and *air pollution* both act as subjects and *it* is semantically replaceable by *air pollution*: *Air*

pollution is one of the most important problem because it can destroy human's health. Therefore, the referent for *it* is identified as *air pollution*.

Stage 2 Identifying key concepts

Having prepared the texts, the next stage is to identify the key concepts in each text. A concept is taken as being a psychological construct which represents some entity in the world and which is represented in discourse through words. Thus, although it is tempting to equate concepts with words, it should be remembered that a concept is a psychological construct whereas a word is a linguistic phenomenon. Because of this difference, we will consider both nouns and noun phrases as potential concepts.

Another aspect of concepts that is important in the analysis is that they serve Halliday's (1970) ideational metafunction of language. Thus, in identifying concepts, we can ignore function words (Read, 2000), words which serve primarily interactional purposes (e.g. *I think*), and words which serve primarily textual purposes such as those used to describe textual schemata (e.g. *problem, reason, cause*).

After identifying those nouns and noun phrases representing concepts in a text, these are then grouped into the concepts they represent by looking for repetition and paraphrase following the algorithm for identifying reiterated items suggested by Hoey (1991). From the list of all possible concepts in a text, the key concepts are identified based on two principles. First, key concepts are frequent (Scott, 1997). In this study, a minimum threshold of 2 occurrences was set for a concept to be considered a key concept. Second, key concepts are salient (Scott, 2000), and thus concepts in titles or concepts highlighted by, say, underlining are counted as key concepts. The key concepts in text 1 are indicated by superscript numbers.

Stage 3 Identifying relationships between key concepts

Having identified the key concepts, we now need to find a way of identifying their distance in semantic space from each other, and this is taken as meaning how closely related they are to each other. There are two key relationships that can be used to link concepts. First and most important is McCarthy's (1988) inclusion relationship, which covers a range of superordinate-subordinate relationships such as hyponymy, meronymy, possessor-possessed, and entity-characteristic. The second relationship is cause-effect.

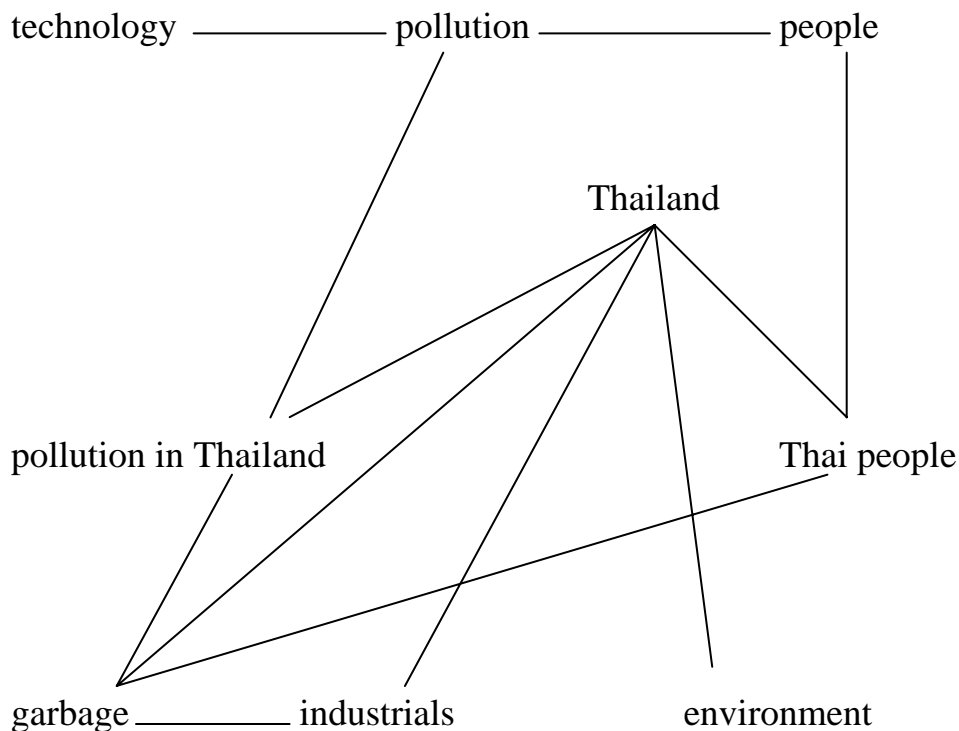
In text 1, examples of such relations include the following:

- *Pollution in Thailand* is semantically a subordinate of, and thus included in, the broader concept *pollution*.
- According to the text, *technology* causes *pollution* and thus there is a cause-effect relation between these two concepts.

Stage 4 Linking the relationships into a hierarchy

Given that inclusion implies a superordinate-subordinate relationship, any set of inclusion relations forms a hierarchy. Cause-effect relations, however, do not form any particular relationship within a hierarchy. We will therefore treat cause-effect relations as forming a link between two concepts without affecting their placement in the hierarchy. Combining the relations between key concepts in text 1 in this way, we can create a hierarchy of key concepts for text 1, and this is shown in Figure 1.

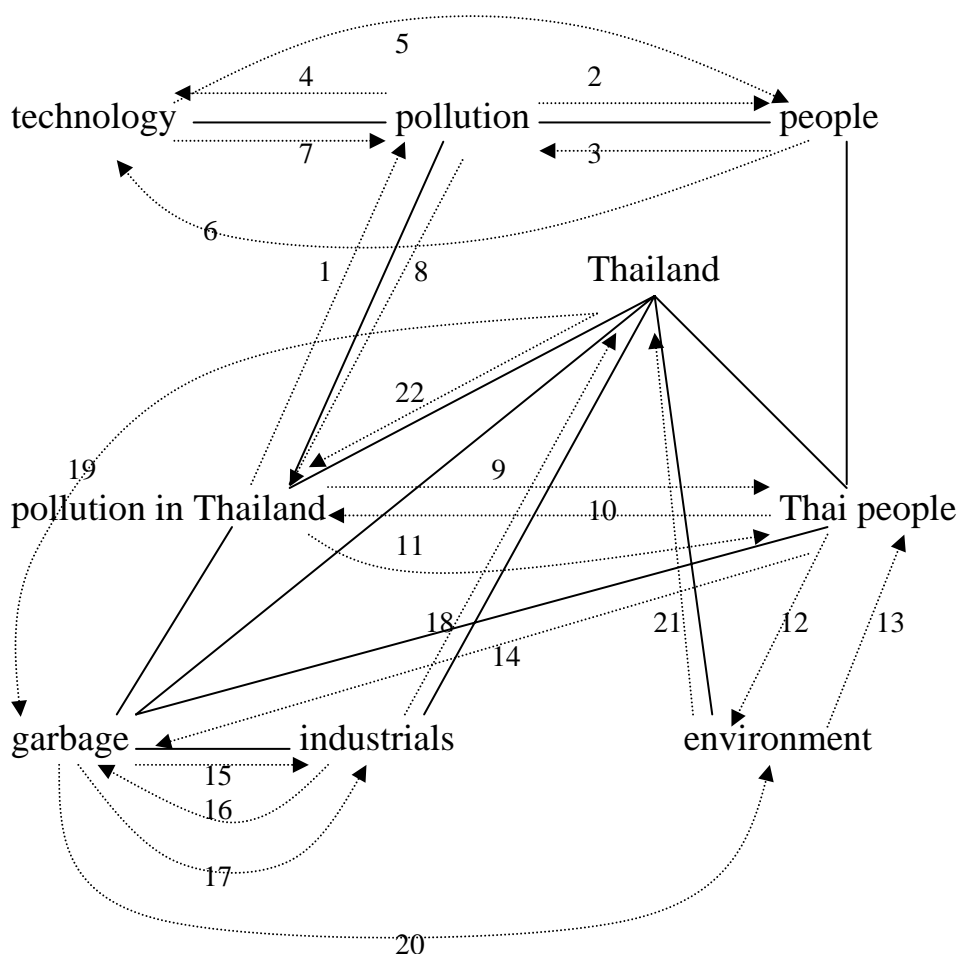
Figure 1 A hierarchy of the key concepts for Text 1



Stage 5 Mapping the discourse onto the hierarchy

Having generated a hierarchy of key concepts for a text, we can then map the moves between the concepts in the discourse onto the hierarchy. For example, the first key concept in text 1 is the title, *pollution in Thailand*, and the second is *pollution*, the first word of the text itself. The first move in the discourse therefore is from *pollution in Thailand* to *pollution*. Mapping all of the moves onto the hierarchy results in Figure 2.

Figure 2 Moves between concepts mapped onto the hierarchy for text 1



Stage 6 Identifying topics and measuring coherence

Although Figure 2 may look somewhat confusing, it does allow us to tentatively identify topics, topic progression and the amount of coherence in text 1. It should be remembered that all of these are subjective facets of discourse involving interaction between the reader and the text, so our identifications are only provisional.

Following de Beaugrande and Dressler's (1981) guideline that density of linkage is indicative of topics, we can see that the key concepts most frequently linked by moves in the discourse are *pollution*, *pollution in Thailand*, *Thai people* and *garbage*. Of these, *pollution* only occurs in the first 8 moves in the discourse. It therefore seems that we can identify *pollution* as the topic of the first six T-units of text 1, which would also appear to exhibit topic maintenance. The other three frequently linked key concepts appear throughout the rest of the text, but the second part of

the text both starts and finishes with *pollution in Thailand* suggesting that this concept should be given priority in identifying the topic. This second part of text 1, however, is less clearly centred around a single topic than the first part.

More pertinently for our purposes, Figure 2 also allows us to generate some quantitative measures of coherence for text 1. Each of the moves between key concepts can be assigned a distance depending on how closely linked the two concepts are in the hierarchy. Where there is a direct link (such as between *pollution in Thailand* and *pollution* in the first move), we can assign a value of 1; where there is no direct link, such as between *people* and *technology* in the fifth move, we can follow the shortest route in the hierarchy to assign a value (in the case of the fifth move, this would be 2). Distances of moves between concepts are important as more distant moves indicate problems with coherence. While a move between two concepts separated by a distance of 1 may be easy for a reader to follow, greater distances may involve a greater processing load for the reader, and therefore a distance of, say, 3 may make it very difficult for the reader to understand how the succeeding concepts relate.

Assigning distance values to moves allows us to create three potential measures of coherence for a text:

1. Average distance of moves in a text. For text 1, all moves have a value of either 1 or 2 and the average distance of moves is 1.36. A lower average distance suggests a more coherent text since there are fewer moves between distant key concepts.
2. Percentage of coherence breaks. Moves with a high distance value are likely to indicate distinct shifts between topics in the discourse or coherence breaks (Wikborg, 1990). For the purposes of this study, we will set a threshold of a distance of 3 to indicate a coherence break (meaning that none appear in text 1). The greater the proportion of coherence breaks in a text, the less coherent the text. This facet of coherence can be measured by examining the percentage of moves that represent coherence breaks.
3. Number of moves/10 T-units. A text with fewer moves between key concepts within a certain length may be easier for a reader to follow and therefore exhibit a greater level of coherence. This characteristic of a text can be measured by looking at the number of moves between key concepts within a length of discourse. Text 1, for example, consists of 16 T-units and contains 22 moves, giving a ratio of number of moves to 10 T-units of 13.75.

Having seen that topic-based analysis can be applied to writing and that it can produce several quantitative measures of coherence, we are now in a position to examine whether these measures of coherence have any reliability and validity. To do this, we will examine the coherence in a selection of students' writing and compare the scores generated by the topic-based analysis with scores generated by experienced teachers.

The data

The texts analysed in this study were undergraduate chemistry students' written assignments on a task-based course at King Mongkut's University of Technology Thonburi, a respected government university in Thailand. These students had been learning English at school for between 6 and 9 years, but were generally at a lower-intermediate proficiency level. While reasonably competent at reading, their productive skills, including writing, were weak. Thirty-four students were assigned to write on the topic of *Pollution*, since it was assumed that the students were familiar with the topic. Of the finished assignments, 6 were disregarded from further analysis as the samples were very short (less than 50 words). Therefore, 28 student assignments formed the data for this study. These were termed texts 1 to 28. Text 1 above is a typical example of these assignments.

The procedures

The 28 texts were analysed following the procedures of topic-based analysis to generate the three measures of coherence described above.

Five experienced teachers (named teachers A to E) were also asked to rate the texts in two ways. These five teachers had all taught courses similar to the one the students were taking, and all of them had experience of both teaching writing and marking written assignments.

To obtain teacher ratings of the assignments, copies of the 28 assignments were distributed to the five teachers who were asked to return them within a week having given two ratings. Firstly, they were asked to give a mark for coherence out of ten for each of the texts. Secondly, the teachers were asked to indicate where they perceived coherence breaks in the texts. No explicit training in how to rate the texts was given, since in the situations where they teach, the five teachers are often asked to mark organisation or coherence without any training and without any clear descriptors. Furthermore, there are no generally accepted guidelines for coherence available, and any guidelines given could be designed to bias the results. However, to avoid problems with technical terms, coherence breaks were described as jumps or breaks between ideas. All five teachers

stated that they were familiar with the concept of jumps between ideas and thus of coherence breaks.

The reliability of the teachers' marks for coherence was checked by asking the teachers to mark 5 randomly chosen texts twice. The pairs of scores were then compared using Pearson's correlation coefficient to measure mark/re-mark reliability (Heaton, 1975). While four of the teachers showed reliability correlations greater than 0.40, the correlation for teacher A was -0.08 suggesting that her marking was unreliable. She was therefore excluded from the analysis.

To assess the reliability of topic-based analysis, the average distance of moves and the number of moves/10 T-units were compared against the teachers' scores. Furthermore, the percentage of coherence breaks identified through the topic-based analysis was compared against teachers' identification of coherence breaks.

In addition to indicating the reliability of topic-based analysis as a measure of coherence, the correlation between scores from the topic-based analysis and teachers' marks can also give some indication of the validity of the analysis.

Findings

Measures of coherence from the topic-based analysis

The application of topic-based analysis to the 28 texts produced three different measures of coherence for each text: average distance of moves, percentage of coherence breaks, and number of moves/10 T-units. The scores are given in Table 1.

Table 1 Coherence scores generated by topic-based analysis

Text	Average distance of moves	Percentage of coherence breaks	Number of moves/ 10 T-units
1	1.36	0.00	13.75
2	1.37	3.57	20.00
3	1.17	5.00	12.86
4	1.72	9.09	24.61
5	1.35	0.00	13.33
6	1.43	3.03	20.62
7	1.39	0.00	15.33
8	1.56	12.50	24.61
9	1.23	0.00	16.84
10	1.59	4.55	18.12
11	1.32	2.44	20.50
12	1.50	2.44	12.38
13	1.52	12.00	20.83
14	1.86	7.69	17.50
15	1.38	4.76	11.67
16	1.58	11.54	15.56
17	1.61	10.00	13.21
18	1.51	3.16	18.00
19	1.46	1.74	18.37
20	1.52	3.12	18.26
21	1.52	0.00	12.27
22	1.64	10.26	26.00
23	1.62	12.12	16.10
24	1.38	2.78	11.92
25	1.57	7.50	20.00
26	1.37	0.00	19.23
27	1.80	7.61	14.57
28	1.31	0.00	17.33
All texts	1.49	4.73	17.29

To check the reliability of the three ways of converting findings from topic-based analysis into scores, we can compare the sets of scores using Pearson's correlation coefficient. The results are given in Table 2.

Table 2 Correlations between the three measures of coherence from topic-based analysis

	No. of moves/10 T-units	% of coherence breaks
Average distance of moves	$r = 0.25$ (not significant)	$r = 0.39; p < 0.05$
% of coherence breaks	$r = 0.16$ (not significant)	

From Table 2, we can see that the average distance of moves and the percentage of coherence breaks are more correlated to each other than either is to the number of moves per 10 T-units, suggesting that the first two measures are more closely related.

Teachers' marks for coherence

The marks given for coherence for each text by the four teachers (teachers B to E) are shown in Table 3.

Table 3 Teachers' marks for coherence

Text	Teacher B	Teacher C	Teacher D	Teacher E	Mean mark
1	10.0	9.0	7.0	9.0	8.75
2	9.0	8.0	4.0	7.0	7.00
3	8.0	7.0	6.0	7.0	7.00
4	9.0	9.0	7.0	8.0	8.25
5	9.0	7.0	6.0	5.0	6.75
6	9.0	8.0	6.0	10.0	8.25
7	8.0	8.0	7.0	5.0	7.00
8	9.0	8.0	6.0	9.0	8.00
9	8.0	9.0	9.0	10.0	9.00
10	9.0	8.0	8.0	5.0	7.50
11	10.0	9.0	7.0	8.0	8.50
12	8.0	7.0	6.0	6.0	6.75
13	10.0	8.0	6.0	7.0	7.75
14	10.0	7.0	6.0	9.0	8.00
15	8.0	6.0	6.0	6.0	6.50
16	9.0	8.0	5.0	7.0	7.25
17	9.0	7.0	6.0	7.0	7.25
18	8.0	7.0	7.0	9.0	7.75
19	9.0	7.0	5.0	7.5	7.12
20	9.0	8.0	7.0	8.0	8.00
21	10.0	9.0	9.0	9.0	9.25
22	10.0	9.0	9.0	8.0	9.00
23	9.0	9.0	9.0	7.0	8.50
24	6.0	8.0	6.0	4.0	6.00
25	9.0	8.0	7.0	7.0	7.75
26	8.0	8.0	6.0	9.0	7.75
27	6.0	7.0	6.0	9.0	7.00
28	n.a.	7.0	5.0	n.a.	6.00
Mean	8.75	7.93	6.71	7.48	7.66

Note: Teachers B and E refused to give marks for coherence to Student 28 because they could not understand the language. Student 28 is therefore discounted from further analysis.

Table 3 reveals a reasonable amount of variation in teachers' marks for coherence. We can check the amount of variation by calculating the correlation between the marks of each pair of teachers. These are shown in Table 4.

Table 4 Correlations between teachers' marks for coherence

	Teacher B	Teacher C	Teacher D
Teacher E	$r = 0.67; p < 0.001$	$r = 0.33$ (not significant)	$r = 0.30$ (not significant)
Teacher D	$r = 0.31$ (not significant)	$r = 0.62; p < 0.001$	
Teacher C	$r = 0.37; p < 0.05$		

The Pearson's correlation coefficients given in Table 4 suggest general agreement between teachers, although this agreement is not always significant.

Comparison of topic-based analysis and teachers' marks for coherence

As we did when we compared the different measures of coherence from topic-based analysis and the different marks given by teachers, we can compare the measures derived from topic-based analysis with the teachers' marks to see any correlations between the two. Table 5 shows the results of this.

Table 5 Correlations between measures of coherence derived from topic-based analysis and teachers' marks for coherence

	Teacher B	Teacher C	Teacher D	Teacher E	Mean score
Mean distance of moves	$r = 0.14$ (not significant)	$r = -0.03$ (not significant)	$r = 0.08$ (not significant)	$r = 0.16$ (not significant)	$r = 0.14$ (not significant)
Percentage of coherence breaks	$r = 0.18$ (not significant)	$r = 0.02$ (not significant)	$r = -0.04$ (not significant)	$r = 0.02$ (not significant)	$r = 0.06$ (not significant)
No. of moves/10 T-units	$r = 0.42;$ $p < 0.05$	$r = 0.43$ $p < 0.05$	$r = 0.10$ (not significant)	$r = 0.39$ $p < 0.05$	$r = 0.46$ $p < 0.05$

Table 5 suggests little relationship between mean distance of moves and percentage of coherence breaks, on the one hand, and the teachers' scores for coherence, on the other. However, the number of moves per 10 T-units correlates with the scores of three of the four teachers and with the mean score for all four teachers, suggesting that, of the three measures of coherence, density of concepts in the assignments is the most likely to be a consideration in the teachers' ratings.

Comparison of topic-based analysis and teachers' identification of coherence breaks

A second approach to measuring coherence is to identify those points in the text where a break in coherence occurs. Such points are identified by

topic-based analysis as those points where a move between concepts involves a distance of 3 or more links in a hierarchy. The teachers were also asked to identify points where an abrupt shift in ideas seemed to occur. The numbers of coherence breaks identified by each teacher are given in Table 6.

Table 6 Number of coherence breaks identified by each teacher

Teacher	B	C	D	E
No. of breaks identified	22	29	131	24

Table 6 demonstrates that teacher D was using a different basis for identifying coherence breaks from the other three teachers. For example, unlike the other teachers, teacher D identified all paragraph breaks in the assignments as coherence breaks regardless of whether the concepts either side of the paragraph break were the same. For this reason, those points at which teacher D is the only teacher to identify a break will not be considered in the analysis, leaving a total of 17 breaks identified by teacher D.

Topic-based analysis identified 21 coherence breaks, which is similar to the numbers identified by most of the teachers. Even though the total numbers of coherence breaks identified are similar, the points at which they were identified show some variation.

There are some points at which all teachers and topic-based analysis agree that there is a coherence break. For example, the move from *tree* to *computer* in the following is consistently identified as a coherence break.

Poison gas is can't reduce by tree. / Computer was develop for help people in many way such as calculate or playing game.

There are also some points at which there is very little agreement concerning the location of coherence breaks. To compare the amount of agreement between the four teachers and between the teachers and topic-based analysis, we can look at the points where two or more teachers agree in the identification of a coherence break and we can also examine the points at which one or more teachers and topic-based analysis agree. The figures for these are shown in Table 7. To clarify the table, we can see that there are 9 points at which 4 agree. Some of these are points

where the four teachers (but not topic-based analysis) identify a coherence break, and some are points where three of the teachers and topic-based analysis identify a break.

Table 7 Comparison of identification of coherence breaks by teachers and by topic-based analysis

	Total no.	No. at which teacher B agrees (N=22)	No. at which teacher C agrees (N=29)	No. at which teacher D agrees (N=22)	No. at which teacher E agrees (N=24)	No. at which topic-based analysis agrees (N=21)
Points at which all 5 agree	2	2	2	2	2	2
Points at which 4 agree	9	8	9	5	9	5
Points at which 3 agree	8	2	4	5	8	5

Table 7 shows that the number and location of coherence breaks identified by topic-based analysis are comparable to those identified by teachers. This suggests that, to the limited extent that inter-teacher identification of coherence breaks is reliable, topic-based analysis is as reliable as a teacher. However, given the amount of variation between teachers in identifying coherence breaks, we should perhaps conclude that teachers and topic-based analysis are similarly unreliable in identifying coherence breaks.

Discussion

From the findings, three notable similarities between the results of topic-based analysis and the teachers' evaluations of the texts emerge:

- The number of moves per 10 T-units correlates closely with the teachers' scores for coherence.
- The total numbers of coherence breaks identified by topic-based analysis and most teachers are similar.
- The levels of agreement concerning the location of coherence breaks for topic-based analysis and teachers are similar.

Given the subjective nature of coherence, it is unrealistic to expect any single rater or method to be entirely reliable. These similarities between

topic-based analysis and the teachers' evaluation therefore should not be stressed too heavily. Nevertheless, they do allow us to make two tentative conclusions.

First, number of moves per 10 T-units is indicative of the density of concepts in a text, and the correlations for this measurement suggest that the teachers are more likely to rate denser texts as more coherent. This is a point warranting further research. If it turns out that density of concepts is an important consideration in coherence, then it should be noted that, of the three topic-based analysis measurements made in this study, the number of moves per 10 T-units is the most straightforward to calculate. The number of moves per 10 T-units can be calculated after the first two of the six stages of topic-based analysis described above, since there is no need to construct hierarchies to obtain this measurement. As well as being less time-consuming, measuring the number of moves per 10 T-units is, at least in theory, programmable on computer. Initially T-units could be identified using a parsing program, ellipsis and referents could be identified using programs based on centering theory (see e.g. Grosz et al., 1995; Walker et al., 1998), a thesaurus could be used to identify paraphrases of the same concept, and frequency count software could then identify the key concepts. If the number of moves per 10 T-units is a key basis for teachers' evaluations of coherence, such a computer program could enable reasonably reliable automated marking of coherence. While theoretically possible, it should be noted that current computer capabilities render such a program unlikely in the foreseeable future.

Second, the similarities in number and placement of coherence breaks suggest that coherence breaks may be a more promising area to focus on in further investigations than coherence per se. In addition, being able to identify coherence breaks may have immediate implications for the teaching of writing. If students can identify where they have made coherence breaks in writing and can repair these breaks, the quality of their writing should improve.

In addition to the three points of relative similarity between topic-based analysis and teachers identified above, there are two aspects for which the findings from topic-based analysis do not relate closely to teachers' evaluations:

- Mean distance of moves does not correlate with teachers' scores for coherence.
- Percentage of coherence breaks within any given text does not correlate with teachers' scores for coherence.

These differences suggest that teachers are not considering distance of moves or frequency of coherence breaks when they assign marks for coherence.

Overall therefore, the findings of this study are mixed. Although teachers may give some aspects of coherence (density of concepts) greater weight when marking coherence than other aspects (mean distance of moves, percentage of coherence breaks), the generalisability of these findings is unclear given the size of the study. Furthermore, the lack of reliability either between teachers or between the teachers and the analysis further weakens our ability to draw conclusions. Some of this lack of reliability may be due to the inherently subjective nature of coherence, but some may also be due to the lack of clear guidelines for teachers concerning how to mark coherence. Although we believe that our decision not to provide guidelines was valid for the purposes of this study, it would be interesting to see the effects of different sets of guidelines on teachers' marking and on the reliability of the marks given.

Conclusion

Conducting topic-based analysis is a complex and laborious process, perhaps too complex and laborious to be worth considering for assessing writing. From this paper, however, we believe that five possible reasons exist for considering the use of topic-based analysis to assess writing:

1. While topic-based analysis may be too complex for teachers to use to assess writing, it may be of value to researchers who wish to attempt to measure the coherence of texts as part of their research.
2. Topic-based analysis could be used as a basis for moderating scoring for coherence with a moderator and markers discussing the analysis in the hope of achieving similar criteria for assigning marks for coherence.
3. If unguided teacher marking of coherence lacks reliability and validity and teachers work in situations where neither moderation nor guidelines for marking are common, training in topic-based analysis may provide a way to raise teachers' awareness of issues of relevance to coherence.
4. Instead of conducting a rigorous topic-based analysis, a looser, more intuitive analysis could still be of value without proving too laborious for teachers who wish to focus in detail on coherence.
5. The correlation between the number of moves per 10 T-units and teachers' marks suggests that this measure, which is easily obtainable, could be used to guide the assessment of coherence.

References

- Bell, R. T. (1991) *Translation and Translating: Theory and Practice*. London: Longman.
- Cohen, A. D. (1994) *Assessing Language Ability in the Classroom*, second edition. Boston: Heinle & Heinle.
- Connor, U. and Farmer, M. (1990) The teaching of topical structure analysis as a revision strategy for ESL writers. In Kroll, B. (ed.) (1990) *Second Language Writing: Research Insights for the Classroom*. Cambridge: Cambridge University Press. pp. 126-139.
- Crow, B. K. (1983) Topic shifts in couples' conversations. In Craig, R. T. and Tracy, K. (eds.) (1983) *Conversational Coherence: Form, Structure and Strategy*. Beverley Hills, CA: Sage Publications. pp. 136-156.
- Daneš, F. (1974) Functional sentence perspective and the organization of the text. In Daneš, F. (ed.) (1974) *Papers on Functional Sentence Perspective*. Prague/The Hague: Academia/Mouton. pp. 106-128.
- De Beaugrande, R. and Dressler, W. (1981) *Introduction to Text Linguistics*. London: Longman.
- Fries, P. H. (1994) On theme, rheme and discourse goals. In Coulthard, M. (ed.) (1994) *Advances in Written Text Analysis*. London: Routledge. pp. 229-249.
- Grosz, B. J., Joshi, A. K. and Weinstein, S. (1995) Centering: a framework for modelling the local coherence of discourse. *Computational Linguistics* vol. 21 no. 2 pp. 203-225.
- Halliday, M. A. K. (1970) Language structure and language function. In Lyons, J. (ed.) (1970) *New Horizons in Linguistics*. Harmondsworth: Penguin. pp. 140-165.
- Heaton, J. B. (1975) *Writing English Language Tests*. London: Longman.
- Hoey, M. (1991) *Patterns of Lexis in Text*. Oxford: Oxford University Press.
- Hughes, A. (2003) *Testing for Language Teachers*, 2nd edition. Cambridge: Cambridge University Press.
- Hunt, K. (1970) Syntactic maturity in school children and adults. *Monographs of the Society for Research in Child Development* serial no. 134 vol. 35.
- Lautamatti, L. (1978) Observations on the development of the topic in simplified discourse. In Connor, U. and Kaplan, R. B. (eds.) (1987) *Writing across Languages: Analysis of L2 Text*. Rowley, MA: Addison-Wesley. pp. 87-113.
- Lautamatti, L. (1990) Coherence in spoken and written discourse. In Connor, U. and Johns, A. M. (eds.) (1990) *Coherence in Writing*:

- Research and Pedagogical Perspectives*. Alexandria, VA: TESOL, Inc. pp. 29-39.
- Lee, I. (2002) Teaching coherence to ESL students: a classroom inquiry. *Journal of Second Language Writing* vol. 11 no. 2 pp. 135-159.
- Mann, W. C. and Thompson, S. A. (1988) Rhetorical Structure Theory: toward a functional theory of text organization. *Text* vol. 8 no. 3 pp. 243-281.
- McCarthy, M. (1988) Some vocabulary patterns in conversation. In Carter, R. and McCarthy, M. (eds.) *Vocabulary and Language Teaching*. London: Longman. pp. 181-200.
- Read, J. (2000) *Assessing Vocabulary*. Cambridge: Cambridge University Press.
- Redeker, G. (1990) Ideational and pragmatic markers of discourse structure. *Journal of Pragmatics* vol. 14 no. 3 pp. 367-381.
- Schneider, M. and Connor, U. (1990) Analysing topical structure in ESL essays: Not all topics are equal. *Studies in Second Language Acquisition* vol. 12 no. 4 pp. 411-427.
- Scott, M. (1997) PC analysis of key words - and key key words. *System* vol. 25 no. 2 pp. 233-245.
- Scott, M. (2000) Focusing on the text and its key words. In Burnard, L. and McEnery, T. (eds.) (2000) *Rethinking Language Pedagogy from a Corpus Perspective: Papers from the Third International Conference on Teaching and Language Corpora*. Amsterdam: Peter Lang. pp. 103-121.
- Sotillo, C. (1999) Reading texts: processing discourse. Available at: <http://www.hcrc.ed.ac.uk/~cathy/discourse.htm>.
- Stubbs, M. (1983) *Discourse Analysis: The Sociolinguistic Analysis of Natural Language*. Oxford: Blackwell.
- Walker, M. A., Joshi, A. K. and Prince, E. F. (1998) Centering in naturally-occurring discourse: an overview. In Walker, M. A., Joshi, A. K. and Prince, E. F. (Eds.) *Centering Theory in Discourse*. Oxford: Clarendon Press. pp.1-28.
- Watson Todd, R. (1997) Textual patterns in teachers' eliciting. *RELC Journal* vol. 28 no. 1 pp. 1-14.
- Watson Todd, R. (1998) Topic-based analysis of classroom discourse. *System* vol. 26 no. 3 pp. 303-318.
- Watson Todd, R. (2003) *Topics in Classroom Discourse*. Unpublished PhD dissertation, University of Liverpool.
- Weigle, S. C. (2002) *Assessing Writing*. Cambridge: Cambridge University Press.
- Weir, C. (1993) *Understanding and Developing Language Tests*. New York: Prentice Hall.

- Weir, C. J. (1990) *Communicative Language Testing*. New York: Prentice Hall.
- Wikborg, E. (1990) Types of coherence breaks in Swedish student writing: Misleading paragraph division. In Connor, U. and Johns, A. M. (eds.) (1990) *Coherence in Writing: Research and Pedagogical Perspectives*. Alexandria, VA: TESOL, Inc. pp. 131-148.
- Yule, G. (1996) *The Study of Language*, second edition. Cambridge: Cambridge University Press.