ASp
la revue du GERAS

4 | 1994
Actes de l'atelier New Technologies du congrès
ESSE/2

Using computers to monitor student performance in essay-writing

Jean Vaché

Electronic version

URL: <http://journals.openedition.org/asp/4132>
DOI: 10.4000/asp.4132
ISSN: 2108-6354

Publisher

Groupe d'étude et de recherche en anglais de spécialité

Printed version

Date of publication: 1 July 1994
Number of pages: 33-42
ISSN: 1246-8185

Electronic reference

Jean Vaché, « Using computers to monitor student performance in essay-writing », *ASp* [Online],
4 | 1994, Online since 19 January 2014, connection on 19 April 2019. URL : <http://journals.openedition.org/asp/4132> ; DOI : 10.4000/asp.4132

This text was automatically generated on 19 April 2019.

Tous droits réservés

Using computers to monitor student performance in essay-writing

Jean Vaché

- 1 I work in the Department of English at the Université Paul Valéry (UPV), Montpellier and I teach English Literature. For years I have been trying to make available to French students some of the computer-based techniques I have seen used at various American universities.
- 2 This particular programme – the one that uses computers in the teaching of literature – has been in operation for several years now. Five years ago when I embarked upon what was then an experiment, I simply purchased a site license for a commercial word processor and asked the students to type their essays on the Macintosh machines to be found in the “Pavillon Informatique”, which houses the general-facility computers. This was then regarded as an oddity by my colleagues as my students were at that time the only non computer-science students to use the facilities.
- 3 Of course, today everything is different. Almost every student at UPV wants to have access to a computer. Now I have a special computer classroom, with an Ethernet Local Area Network, a laser printer and we use a sophisticated piece of software, the Daedalus Integrated Writing Environment, which will be described in detail below.

Materials and methods

- 4 I will first explain how literature courses for English majors work at my university. Let us bear in mind the fact that we are dealing with ESL students (English as a second language). There are two types of literature classes: a ‘directed study class’ which is taught by a member of staff who lectures for two hours weekly on the required curriculum to a group of approximately 40 students, and a ‘practical studies’ class of 15 students, taught by an assistant or *lecteur*, usually a graduate student from Birmingham or Austin, which meets one hour per week. The curriculum itself is a survey of English and American literature ranging from Chaucer to the 20th century. Part of the course requirement is that the student should, during the school year, submit several papers

based on topics provided in advance by their teachers. Before the “invention” of the computer classroom at my university, our students were offered no specific training in English composition; they were somehow expected to transpose the skills they had acquired in French composition (the French academic ‘dissertation’) to English and that was how they were supposed to learn the art of the English essay.

- 5 My idea was to capitalize on the use of the computers to change radically the nature of the ‘practical studies’ classes and create real writing workshops, where students could be shown how to write an essay and develop their own skills. Relying on the method described by William Wresch (1984), in which writing is considered as a process rather than a product, I devised the following pedagogical method:
 - Students write the first draft of their papers on the computer, submitting it in print to their *lecteurs* (literally, readers) for correction.
 - The *lecteurs* hand back the papers with suggestions for improving style and coherence, as well as the more technical aspects like spelling and grammar.
 - The students then revise the electronic form of their papers before turning the final draft in to their ‘directed study’ teacher, the idea being that by the time an essay reaches the teacher, it should, theoretically, have been “debugged”.
- 6 More than half the ‘practical studies’ (*lecteurs*) classes meet in our special computer classroom. The rest continue to work in a more traditional environment, that is to say, with paper and pen. It is obviously easier for the computer class students using electronic writing techniques to manipulate their papers than it is for the students in the traditional groups. Moreover, teachers do not usually try to conduct a pedagogical study based on a corpus made of indecipherable hand-written papers; they know it is a hopeless task. But the advent of electronic texts in the new classroom is changing all that. This paper is therefore an attempt to explore some new evaluation techniques offered by modern technology. The idea is to use the available data of students’ written productions to devise a method of evaluation based on statistics. The method used tentatively this year will be described in the second part of this paper. A word of explanation on the software used at Montpellier is in order first.
- 7 The Daedalus Integrated Writing Environment (DIWE) is a fully integrated computer-based collaborative system designed for the English classroom. Through five programs, it helps teachers to explore the possibilities of text-sharing pedagogy:
 - “Write” is a simplified word processor that is also the central text editor throughout the other programs. It has a built-in spelling checker and a concordance maker.
 - “Mail” is a simple electronic a-synchronous mail program.
 - “Invent” is an invention heuristic activity that draws would-be “writers” into preliminary self-exploration of their knowledge about topics.
 - “Respond” is a revision heuristic activity that guides the user in responding to paper drafts through a series of evaluative prompts.
 - “Interchange” is a real-time (synchronous) conferencing program in which participants compose messages privately and then send them to all the other members of the discussion group.
- 8 My intention here is not to give a full description of the Daedalus system’s possibilities nor of the process and collaborative approach which we have implemented. My aim is to concentrate on the end-product, the students’ textual productions. Last year, our students’ training in essay-writing took them through three cycles of the revision/

rewriting method described earlier. Having only decided to conduct the present quantitative study in March, the results which I will now present are highly tentative and provisional. Here is a brief description of the method I followed. First, I asked a hundred students who attended the computer classes to retrieve the very first computer-based draft of the first essay they wrote in November/December 1992; only half of them were able to unearth it (some had overwritten their first draft when revising it, others had lost their diskettes, etc.). Then, in May I collected the final draft of their last essay of the year. All essays were of course in electronic form. This was the corpus I worked with (a collection of approximately 100 individual texts). From a purely statistical point of view, the numbers involved are perhaps not very large; however, the total number of words is almost a hundred thousand and computing the results took many hours of work.

- 9 My first decision was to get rid of all the literary quotations we normally encourage our students to use in their essays. If I had not done so, my statistics would have been distorted. This was probably the most arduous task of all. No existing program could be used to erase all quotations automatically. And even if I had written one, how could it have discriminated between indented longer quotations, and the text itself? This meant that I had to read and amend individually more than a hundred essays, each several pages long, and delete all text that was not strictly the student's own. I mercilessly cut out hundreds of Shakespearean, Miltonian, Swiftian, Dickensian, and other Thoreau-esque quotations as a preliminary step to establishing a legitimate corpus.
- 10 The next step was the choice of a concordance tool. After considering professional concordance programmes like the Oxford Concordance Program (OCP) or Tact, I decided to rely on a very simple tool provided with the DIWE, a mini-concordance that works within the program, by-passing the text-only form used by other concordances. The DIWE concordance is also DIY: it can be used by students to evaluate instantly their own written productions; indeed, they can build a concordance that gives them a list of words used, sorted alphabetically or by frequency, they also get a series of statistics on their prose: total number of characters, of words, of sentences, of questions, of paragraphs; the program calculates the number of unique words (the type words, that is to say the number of different words, as opposed to the token words: the total number of words), and the percentages of characters per word, words per sentence, sentences per paragraph. One of the most interesting figures provided automatically is the ratio of type to token words (unique/total ratio) which gives a fairly reliable estimate of vocabulary richness.

Results

- 11 Predictably enough, the sheer overall length of the essays increased between November and May:
- 12
 - the shortest paper #1 was only 647 words long; the longest, 2312.
 - the average length of paper #1 was 1120 words.
 - the shortest paper #3 was only 770 words long; the longest, 3439.
 - the average length of paper #3 was 1528 words.
 - the net increase in average length is 408 words, or 26%.
 - two thirds of the students increased their essay-length from essay 1 to essay 3.

- 13 Another observable result is that the length of words slightly increased between the two papers:
- 14 -from 5.79 in the first essay, to 5.86 in the second, a modest 0.07 character increase. (By comparison, the average number of characters per word in this paper is 6.23)
-two-thirds of the students increased their word-length in their essays.
- 15 As the essays grew in size, students wrote more sentences: from an average of 65.13 in essay #1 to an average of 82.72 in essay #3; a net average increase of 17.59 sentences. And the number of words in each sentence also increased slightly, passing from an average of 18.78 to 19.01.
- 16 Considering now the number of paragraphs, they passed from an average of 7.95 in the first paper, to an average of 11.16 in the last. A net average increment of 3.21 paragraphs.
- 17 The most interesting and probably the most telling set of figures yielded by our concordance is the 'unique words' feature. The number of unique words in a given text describes the degree of rigidity or flexibility in word use. Our concordance takes only a few minutes to extract a list of unique words, a task that I would never have embarked upon if I had been obliged to do it manually. Here again, a little more than two-thirds of the students managed to improve their vocabulary range between the beginning and the end of the year. The average number of unique words in essay #1 was 393.35. The number passed to 495.11 in essay #3. The average increment is 101.76 or 18% for the group's performance. It is not possible to examine all individual performances here, but some cases are interesting. For example, one student managed to enlarge her vocabulary in essay #3 while actually writing a shorter essay.
- 18 However, monitoring unique word occurrences in students' papers is not an end in itself. To be interpreted, such results must be read in the context of the papers' actual length. A simple calculation shows that very short samples of language contain a very high ratio of unique occurrences, while the ratio diminishes as samples grow longer. I have therefore borrowed the concept of TTR (or type/token ratio, a measure of vocabulary diversity) from the American psychologist W. Johnson, who devised this tool to study what he called "language behaviour" (Johnson 1944). TTR is the number of different words in a language sample divided by the total number of words in the sample. The results of such a computation show that my students reached a mean ratio of 0.36 in the first, shorter essay-draft in which the number of words averaged 1,120, and an average of only 0.33 in the last, longer essay in which the total number of words averaged 1,528. Explanations for such an apparent drop in vocabulary richness will be offered in the next and last part of this paper.

Discussion

- 19 If it is relatively simple to explain the various increases observed throughout the year at the word, sentence, paragraph and essay levels (better computer-literacy, finer handling of revision techniques, disappearance of writer's block, etc.), it is less easy to explain the sharp drop in TTR in the final essay. What I needed was to compare my results with language samples obtained from Anglophones. In order to gauge the "acceptable" or "desirable" ratio of unique words in language samples taken from English-speaking *locuteurs* against my ESL students' results, I used first the essays written by three English-born Erasmus students who happened to belong to the same class, then, I scanned a 6-

page literary essay written by one of England's finest contemporary novelists, A.S. Byatt (1990), and finally, I scanned a 6-page technical essay written by Locke Carter, American-born and co-author of the DIWE program (Carter 1992).

- 20 Predictably, all three language samples show a higher ratio of TTR. The English students' average was 0.4 for essay #1 and 0.37 for essay #3, with the average total number of words increasing from 1,100 to 1,372. Here again, I discovered a percentage drop in the last essay, but the ratio remained higher (0.37 to 0.33). A.S. Byatt scored 0.38 for a total number of words of 2655, and Locke Carter's technical essay scored 0.33 for a total 2,744 words. But the most useful figures came from a second scanning of the last two language samples. I built a series of concordances of increasingly longer sections of the texts so as to create a sort of scale of comparison for my students' papers. The results are shown in Table 1 and Table 2.

Table 1: A.S. Byatt's "Under a Stronger Sun"

Total Words	Unique Words	TTR
18	17	0.94
33	27	0.81
104	71	0.68
274	167	0.60
453	256	0.56
819	428	0.52
1,222	575	0.47
1,582	726	0.45
2,199	902	0.41
2,655	1,029	0.38

Table 2: Locke Carter's "Computer mediated discourse and student participation"

Total Words	Unique Words	TTR
19	18	0.94
40	33	0.82
80	53	0.65
218	130	0.59

433	225	0.51
821	364	0.44
1,157	485	0.41
1,659	639	0.38
2,223	816	0.36
2,744	926	0.33

- 21 This confirmed my intuition that the longer the language sample, the lower the TTR. But the drop in the curve was decidedly slower in the “literary” sample than in the “technical” one. By merging the two curves, I obtained a sort of “snake”, the upper part showing the “highly desirable” area, the lower one signalling a zone of acceptability. By comparing my students’ papers to the “snake” I was able to determine whether they had reached a threshold of acceptability regardless of the actual length of their papers. The results of such a comparison show that in the case of the first essay, 23% of the students barely managed to reach that threshold, but they wrote shorter papers which made the TTR less meaningful. In the second, longer essay however, only 9% achieved the goal of acceptability (as far as TTR was concerned). This result in itself paints perhaps a more convincing and more realistic picture of ESL students still struggling with written English and the essay genre.
- 22 From a more general point of view, the first piece of criticism that could be levelled at these results is that the number of essays (3) actually processed to the final stage of completion was obviously too small to constitute adequate training for a complex exercise like essay-writing as well as an adequate sampling for a statistical study. If this is true, then surely no valid findings can be drawn from the previous figures. This, added to the author’s lack of competence in statistical matters, should warn us against jumping the gun in our conclusions. However, it may not be unreasonable to present a few observations.
- 23 A first observation on the method used here involves questioning the reliability of the type/token ratio (TTR) as a measure of vocabulary diversity. As I have shown, it needs to be corrected by “length” on the assumption that as a writer proceeds he or she will repeat key vocabulary, quite properly so. Therefore, in addition to the fact that any statistical study must include a general coefficient for length that can be applied to any TTR, any statistical study concerning ESL student productions must include the notion of a “threshold of acceptability”. With this “caveat”, my original intuition is perhaps confirmed: computer-aided instruction in writing does facilitate pedagogical investigations of students’ productions and students’ productions can be monitored at two levels: the individual, and the collective level. Computer-aided instruction in writing makes the monitoring easier of the individual student, who progressively acquires a better command of his or her written productions, and this evolution is perhaps reflected in the improvements we detected at various levels: overall length of text, sentence length, word length, and most of all, probably, word choice. At the collective level, results are even more interesting to watch. If several groups were monitored this way over a

longer period, it might be possible to test not just the quality of our students, but also the quality of our teaching.

- 24 A second observation concerns what is actually monitored in a study like this. You will perhaps agree with me that a limited number of linguistic behaviours or skills can be observed and recorded by machines such as computers, with the help of astute tools like pieces of software which are nothing but extensions of the human mind; this is what I have tried to demonstrate here. But there can be no question that literary skills as such cannot be directly observed by machines. My younger colleagues will never be replaced by robots. I simply wish to show that with the probable future extension of the use of computers to teach writing, new perspectives will be opened, and better tools developed to monitor the writing process and the way we teach it.

BIBLIOGRAPHY

- Berlin, James A. 1982. "A contemporary composition: The major pedagogical theories." *College English* 44, 765-777.
- Byatt, A.S. 1990. "Under a Stronger Sun". *The Independent Magazine* 3 March.
- Carter, Locke. 1992. "Computer mediated discourse and student participation." *Cahiers de l'APLIUT* 45, 66-76.
- Faigley, Lester and Stephen Witte. 1981. "Analysing revision." *College Composition and Communication* 32, 400-414.
- Hillocks, Jr, George. 1982. "The interaction of instruction, teacher comment, and revision in teaching the composing process." *Research in the Teaching of English* 16, 261-278.
- Horner, Winifred B. 1983. *The Present State of Scholarship in Historical and Contemporary Rhetoric*. Columbia and London: University of Missouri Press.
- Johnson, W. 1944. "Studies in language behavior: A program of research." *Psychological Monograph* 56, 1-15.
- Sommers, Nancy. 1980. "Revision strategies of student writers and experienced adult writers." *College Composition and Communication* 31, 378-388.
- Witte, Stephen and Lester Faigley. 1981. "Coherence, cohesion and writing quality." *College Composition and Communication* 32, 189-204.
- Wresh, William (ed.). 1984. *The Computer in Composition Instruction; A Writer's Tool*. Urbana, IL: National Council of Teachers of English.

ABSTRACTS

This paper addresses the problem of evaluating the quality of students' productions by using evaluation methods partially based on quantitative measurements. A corpus has been created consisting of essays on literary topics written over a school year by a group of second-year

students of English in a French university. The essays have been individually parsed and examined by a computer-based concordance that records measurable elements such as total words, unique words, number of sentences, forms of verbs, pronouns, etc. Individual achievement and group performance over a period of time have been recorded in the form of percentages and numbers. Working from these figures, the author will try to answer a series of questions. Can trends be discerned? Can measures like these be considered valid methods of evaluating student performance? And can computers be used to test the quality of our teaching?

Il s'agit ici d'évaluer la qualité de la production écrite des étudiants d'anglais grâce à des méthodes d'évaluation reposant en partie sur des mesures quantitatives. Le corpus est constitué des compositions écrites pendant une année sur des sujets littéraires par un groupe d'étudiants de seconde année. Chaque composition a été traitée par analyseur syntaxique et examinée à partir d'un outil de concordance qui enregistre tous éléments quantifiables tels que nombre total de mots, mots à occurrence unique, nombre de phrases, formes verbales, pronoms, etc. Les résultats tant individuels que collectifs sont exprimés en pourcentages et en valeurs. À partir de ces données, on tente de répondre à un certain nombre de questions : existe-t-il des tendances repérables ? De telles mesures peuvent-elles être considérées comme une méthode valable d'évaluation des étudiants ? Peut-on avoir recours à l'ordinateur pour mesurer la qualité de notre enseignement ?

INDEX

Keywords: concordance, evaluation, writing, writing workshop

Mots-clés: atelier d'écriture, concordance, évaluation, rédaction

AUTHOR

JEAN VACHÉ

Jean Vaché teaches at Université Paul Valéry Montpellier 3. jean.vache@univ-montp3.fr