Statistical Approaches to Philippine Literature

Jose Nilo G. Binongo; M.W.A. Smith

Statistical Approaches to Philippine Literature

Jose Nilo G. Binongo
M. W. A. Smith

Writers, great or otherwise, are alike when they practice their art but are unconcerned about whether or not they are conforming to some "scientific law." Nor do they expect their works to be read like a scientific treatise. The goals of writers and practitioners of science have hardly anything in common. Understandably then, attempts to bridge literature and science can meet with opposition. The reaction of the well-loved English teacher, Mr. John Keating, to the school's poetry textbook *Dead Poets Society* illustrates the rather widespread aversion to what is deemed as the technicalization and dehumanization of literature.

Polemics aside, practitioners of quantitative stylistics (not to mention linguists who have been around long before the former) have, nevertheless, found a wealth of evidence which demonstrates that writers' use of language has an objective, quantifiable dimension. Accordingly, readers and critics alike often talk about how writers alter their language in different works or how they differ from other writers in their consistent use of certain features. For, indeed, the freedom of writers to be creative when composing their works is far from absolute. Notwithstanding the so-called 'poetic license', writers who wish to communicate cannot entirely turn a deaf ear to the basic rules of syntax. Philosophers Mendelsohn and Schwartz's (1987,

260) take note of the fact that "one of the primary functions of grammar is to *disambiguate*: the rules are adopted to ensure that a given sequence of words will have a unique, clear interpretation." Furthermore, as a result of their differing personal, cultural background or socio-politico-economic milieu, writers cannot but have certain predispositions, if not biases. Sometimes, the choices are not conscious. Stanford linguists Traugott and Pratt (1980, 29) elaborate:

> To claim that style is choice is not, of course, to claim that it is always conscious choice. Indeed, if one had to make all phonological, syntactic, semantic, and pragmatic choices consciously, it would take a very long time to say anything at all.

It is thus reasonable to expect writers to exhibit certain recurring grammatical and stylistic preferences, some more perceptible than others, in their works. Indispensable in the understanding of a writer's style, these preferences or tendencies can often be summarized as frequencies of occurrence, the domain of statistics. In the words of Leech and Short (1981, 42):

> The recognition of cohesion and consistency in preference is important: without it, one would scarcely acknowledge a style. To go one stage further: 'consistency' and 'tendency' are most naturally reduced to 'frequency', and so, it appears, the stylistician becomes a statistician.

This article is an attempt to explain primarily the reason for, and manner of, conducting literary-statistical studies. First, reservations of literary scholars are addressed; in so doing, quantification in literary studies is justified. Previous research is sampled to give a flavor of the breadth and purpose of quantitative investigations. Prerequisites for such investigations are then stated. As a further reflection of how quantitative stylistics can contribute to literary scholarship, the final section ends with the note that the relationship between the traditional and computational approaches should be complementary, neither competing nor conflicting. Research in this area has just started in the Philippines.[1] It is hoped that the present discussion will encourage further work.

Van Peer's Philosophical Objections

The main object of quantitative stylistics is to measure *style,* whatever that is. Different scholars have proposed a variety of definitions. One of the early researchers who attempted to quantify style,

Wincenty Lutoslawski (1897, 64), reflecting his stylometric approach, stressed authorial style which, in his words, is "a mark of the identity of an author independently of the contents of his writings." Such a notion is actually traditional: couched in Latin, the expression stil*us virum arguit* (or the French equivalent, *le style, c'est l'homme,* meaning "The style proclaims the man") is familiar.

In their book *Style in Fiction,* Leech and Short (1981, 11) point out that style "has been applied to the way language is used in a particular genre, period, school of writing, or some combination of these: 'epistolary style', 'early eighteenth-century style', 'euphuistic style', 'the style of Victorian novels', etc." Noting that "in practice, writers on style have differed a great deal in their understanding of the subject" they assert that "it would be artificial to limit our understanding of style to one of them, let us say authorial style, and exclude the others."[2] Thus, confronted by such variety, they are content with defining style as simply "the linguistic characteristics of a particular text" (12). This simple definition of style has been accepted for the present article. Notwithstanding its generality, the definition is understood not to include style in speech. This article then imposes a further restriction in that the linguistic characteristics are taken to refer to a literary text.

Quantitative stylistics is often associated with frequency counts, particularly at the *graphological level,* "the level at which letters combine to form orthographic words, and words to form phrases and sentences delimited by punctuation" (Butler 1985a, 12). The distribution of word lengths in terms of letters or of sentence lengths in terms of words, the number of words occurring once *(hapax legomena),* twice *(hapax dislegomena),* and the rate at which a writer generates new words *(type-token ratio)* are some examples of the many quantities a researcher can count and study.

In the field of linguistics, the utility of frequency counts has long been recognized. Frequency dictionaries of the English language, for example, have been assembled—the better known of which are the American-English-based Brown University corpus and the British-English-based Lancaster-Oslo/Bergen (LOB) corpus. "Scholars wanting to compare speech and writing in British English can draw on the computerized London-Lund Corpus (derived from the Survey of English Usage (SEU)) for speech and the LOB Corpus for printed material" (Greenbaum 1991, 83).[3] In the Philippine setting, a former professor of linguistics at De la Salle University, Curtis D. McFarland (1989) has assembled a frequency dictionary of the Philippine national

language.[4] His work, *Frequency Count of Filipino*, is the first of its kind in the country. In the words of Andrew Gonzalez, FSC, former president of the Linguistic Society of the Philippines, the people who would largely benefit from McFarland's work are textbook writers of Filipino who "will be able to produce usage-based materials for beginning readers and thus use functional load as one basis for initial reading materials presentation," future dictionary makers who "will have a principled basis for inclusion of items in the dictionary," and materials writers composing teaching manuals for Filipino as a second language to whom "the word count furnishes a guide for the choice of lexical items which should be used in the sample texts" (Gonzalez 1989, i).

In the literary field, however, the question of relevance comes to the fore. Some literary scholars have gone as far as to question the foundations of quantitative stylistics. In "Quantitative Studies of Literature: A Critique and an Outlook," literary theorist Van Peer (1989, 302) argues:

> The quantitative approach...misses one of language's most fundamental characteristics, i.e., its *transient* nature. The logical flaw involved here lies in an act of *reification*: what in fact is an ongoing and context-bound process, in which meanings are situationally constituted is reified into a distinct and isolated ontological category....That is to say, in the very act of transforming textual qualities into counts, their essentially process-like character is irretrievably lost. Such a loss may be futile if the respective qualities are peripheral to the object under investigation (as is the case, for instance, in physics). Wherever such qualities belong to the very nature of the phenomenon being studied, however, the loss must be deemed detrimental.

Thus, he concludes:

> By turning linguistic qualities of a text into numerical form, then, quantitative studies...undermine the very foundations upon which the study of literature must necessarily be based.

Even prior to Van Peer's article, scholars in quantitative stylistics, cognizant of such objections, have, now and then, reassessed the significance of their work. Burrows (1987, 28–29), for example, in his book, *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method*, considers whether words taken alone can be repositories of meaning. Using the colorless word 'we' as a test case, he observes:

a traditional grammarian would maintain that 'we', like other words, should be regarded as having incipient meaning, in a sort of Aristotelian potentiality, not realized until it is set in context. More recent authorities would maintain that, for any speaker of English, 'we' bears a certain 'context' even before it is brought into use. It is among those words that can open a sentence. It is among those words that can stand as subject to a verb. It is among those words that allude to more than one referent (the speaker being among them) without actually naming those referents.... As soon as it is mentioned, even if it is the opening words of a fresh discourse, 'we' takes on a more immediate meaning by identifying its referents: the bases of identification, not always ambiguous at first, are likely to be predominantly grammatical when other utterances have led on to the 'we,' predominantly social when it initiates a fresh discourse. On either traditional or more recent doctrine, 'we,' taken alone, is not devoid of meaning.

In a more recent essay, "Computers and the Study of Literature," Burrows (1992, 183) answers Van Peer more directly. Taking off from where Van Peer left, Burrows continues the reference to science—this time biology—and argues:

In many but not in all respects, a word or phrase is certainly most meaningful when it is seen in context: the question ... is whether anything further can be learnt by taking a word away for a time from its contextual neighbours and setting it among its semantic or grammatical kinsfolk, other specimens of its own type. In its strength but also in this limitation, Van Peer's argument applies more generally than he concedes to all scientific inquiry into the constituents of living organisms and dynamic systems; and, in stylistics [an acceptable umbrella term for an area of literary studies in which quantitative methods are now prominent], at least, the guinea-pigs survive the experiment, are liberated, and return unblemished to their proper habitat.

To demonstrate this point, Burrows proceeds to survey current research in quantitative stylistics; the studies he mentions (which will be sampled in the next section) provide a variety in scope (as far as literary analysis is concerned) and a profundity in results. Burrows explains that Van Peer's argument applies more to such dynamic systems as "the constituents of living organisms." The present authors, moreover, add that even a system as dynamic and as complex as the human body can afford, though not usually permanently, to lose a part. Temporarily dismembering the whole permits a more scrupulous examination of the part, and when the part is put back to its original place, the body resumes its normal functioning.

In other words, the main issue against Van Peer is that his understanding of the nature of language appears to be incomplete. As Binongo (1995a, 70–71) phrases it when he scrutinizes Nick Joaquin's style from a quantitative perspective,

> admittedly, however, the [quantitative] approach is not without a drawback: there remain important aspects of Joaquin's style that are difficult to quantify, particularly the use of poetic language in prose ...sound, imagery, and especially the figurative which ... resists any attempt at statistical counts and frequencies. But neither can a non-quantitative approach boast of being comprehensive; the aspects of Joaquin's style examined [here: sentence length, vocabulary richness, etc.] ... are those that are best expressed in quantitative terms.

Burrows's own research on what he calls "colourless words" points to the possibility that there are gaps in both contemporary linguistic and literary theory which need to be filled.

"The novelist's medium is language: whatever he does, *qua* novelist, he does in and through language." So writes David Lodge (1966, ix) in *Language of Fiction.* By extracting patterns in a writer's dealing with language, researchers in computational stylistics have pointed out numerous possibilities for a harmonious marriage between stylistic and statistical analysis. Indeed, recent work in computational stylistics has increasingly demonstrated the utility of quantitative approaches to style. Gradually, acceptance has ensued, permitting formulations such as:

> [Style is] the message carried by the frequency distributions and transitional probabilities of its linguistic features, especially as they differ from those of the same features in the language as a whole. (Bloch 1953, 40)

> The style of a text is a function of the aggregate ratios between the frequencies of its phonological, grammatical, and lexical items, and the frequencies of the corresponding items in a contextually related norm. (Enkvist 1964, 28)

Bloch or Enkvist's understanding of style may still be a moot one, but both formulations do illustrate that language is not without a dimension that is very much amenable to quantitative scrutiny.

Samples of Previous Research

The studies examined by Burrows (1992) and those appearing in the periodicals, *Literary and Linguistic Computing, Computers and the*

Humanities, Style, or *TEXT,* show that scholars like Van Peer have failed to discourage further research in computational stylistics. Due to the proliferation of studies, a survey is not possible here. However, a few examples taken from those reviewed by Burrows will illustrate the point. Of course, there are other books containing similar surveys; among them are Rosanne G. Potter's (1989) *Literary Computing and Literary Criticism,* which Burrows himself uses, and Chapter 2 of Butler's (1985a) *Computers in Linguistics.*

The computer-assisted studies of prosody and poetic phonology of Bratley and Ross (1981) and Logan (1982, 1985) are noteworthy. Regarding Logan's study, Burrows (1992, 184) remarks:

> His greatest advance lies in his development of an index of 'metrical complexity,' which relates patterns of metre to the underlying linguistic patterns of normal stress and which thereby admits comparisons between the work of different poets or the methods of different prosodies.

Studies based on concordances constitute another area to which the computer brings obvious benefits. Addressing the problems of narrative reliability in Ford Madox Ford's *The Good Soldier,* Sabol (1989) examines all the instances of the word, "that," "to locate the propositions in which Dowell, the narrator, affirms his knowledge and manifests his doubts" (Burrows 1992, 184). Fortier (1989) "treats of more prominent thematic features—semantic fields like h*appiness, confinement,.* and *night*—and arrives at suggestive and sometimes unexpected contrasts among the four modern French novels" (Burrows 1992, 185). Burrows (185–86) evaluates studies which employ this approach as follows:

> Like Fortier's semantic fields, the image-clusters studied by Ide (1987, 1989a, 1989b) and John B. Smith (1980) are not abstractions, but accumulations of just such revealing particulars as any New Critic would have approved. And, as in the work of the New Critics, the particulars examined by these newer critics are not treated as stable entities. They are regarded, rather, as gathering meaning by mutual association, by their changing roles in the forward movement of a text, and by the shape they derive from and help to confer upon it.

Using highest frequency words as raw data for analysis, Burrows's own investigation reveals very objectively a steady process of change in the English language and "the comparative homogeneity of eight-

eenth-century prose and the comparative diversity of national, re-
gional and personal styles in our own time" (192).

The examination of intrinsic properties of literary works also helps
to understand the development of a writer's style.

> Brainerd (1980), treating only of the changing incidence of personal
> pronouns, is able to establish a chronological grouping of Shakespeare's
> plays. Butler (1981) studies features of the language of Sylvia Plath's
> four volumes of poetry and finds strong evidence of movement from
> a more formal and ornate to an ostensibly plainer but more subtly
> charged style. Burrows (1987a, 1987b) finds evidence, in both dialogue
> and narrative of Jane Austen's novels, of movement from a more
> 'Augustan' to a more fluid style. (Burrows 1992, 186–87)

When researchers find a feature in an author's manner of writing
which does not vary very much with time (that is, irrespective of
the author's growing writing maturity or of decline) and which can
be shown to occur at a reliably different rate from those of other
authors writing in the same genre and period, they may have iden-
tified a stylistic 'signature' of the author. If disputes, often notorious
among its practitioners, constitute evidence of its life, then research
in this area must be in danger of becoming hyperactive; the latest
ones employ a combination of results from neurophysiology and
computer science. The reader may wish to refer to Holmes's (1992)
survey of the field of stylometry, the branch of computational
stylistics which quantifies style primarily "for the purpose of deduc-
ing the authorship of certain works and of determining the chronol-
ogy of an author's works" (Binongo and Smith 1996).

It becomes evident, even from this brief sample of past research,
that statistical approaches to literature continue to play a significant
role, allowing practitioners to remain hopeful for the future of com-
puter-assisted literary analysis. In Burrows's (187) words,

> If, in future, computer-assisted criticism does draw on the legacy of
> structuralism, it seems more likely to gain more from its forerunners,
> especially the Russian Formalists, than from the French and American
> structuralists themselves. If, in the second place, the post-structuralist
> enterprise continues to outlive structuralism, this newer wave of theo-
> rists may find advantage in the capacity of the computer to go over
> the ground in different ways, deconstructing one classificatory model
> and resolving it into another and yet another. In either event, compu-
> ter-assisted criticism is likely to become an ever stronger force in liter-
> ary studies.

Prerequisites of Literary-Statistical Studies

The pivot of the rapid development in the field of quantitative stylistics was, is and probably will continue to be, the computer (Binongo 1995a, 269). In regard to McFarland's *Frequency Count of Filipino*, Brother Gonzalez (1989, i) writes:

> What took the Brown University computational linguists a whole generation to complete with primitive technology was made possible by the work of one person who singly used a modern microcomputer for three years.

"Through its emphasis on ranking and ordering and, by implication, on the logic of classification, the French term, *l'ordinateur*, points to deeper and more general truths than are encompassed by the strictly numerical connotations of our term, 'computer'" (Burrows 1992, 157). Indeed, as research in quantitative stylistics shows, the computer can do more than merely 'compute.' The concept of ordonance implied in the French word for 'computer' "embraces genuine affinities between the methodological bases of literary studies and the computer-assisted collection and analysis of data" (157).

In this respect, the state of affairs in the Philippines is not ideal. Used merely as word processors, the potential of computers in the literature departments of major universities has not been realized. Binongo (1995a, 269) mentioned the establishment of literary and linguistic computing centers in prestigious universities around the world (among the first ones were Oxford University Computing Centre, U.K.; University of Toronto, Canada; Princeton and Rutgers Universities, U.S.A; University of Newcastle, Australia) to respond to the growing need for research in the field of computational stylistics. These centers are equipped not only with a staff well-versed in computing in the humanities but also with large databases of machine-readable versions of major literary works. Moreover, computer programs are readily available in these centers to perform literary and linguistic analyses on the electronically stored texts.

In the Philippines, the establishment of a literary and linguistic computing center may be low in every university's list of priorities. Nevertheless, for computer-assisted literary-statistical studies to flourish, universities would have to upgrade their technological resources. Because some universities still cling to the more-than-a-decade-old IBM PC/XT technology—which, apparently, is unfriendly to recent

technological developments—the picture is gloomy. In the area of storage technology, for example, the use of CD-ROMs to study Shakespeare's Jew or William Faulkner's negro is a phenomenon which has not reached Philippine shores. Moreover, at this time of writing, few computer users have heard of the so-called magneto-optical (MO) disks.[5] Because they are removable (unlike the hard disk) and rewritable (unlike the usual CD ROM), MO disks have posed as a serious threat to the 40-year-old magnetic hard disk technology. They are, in reality, diskettes, but they hold such massive storage capacities as 238 megabytes, 650 megabytes, or even as big as 4.6 gigabytes.[6] Users of the early XT computers were using what were then called double-density (2D) 5.25-inch diskettes, each with a capacity of 360 bytes, less than half of a megabyte. A university facility for literary and linguistic computing would undoubtedly benefit from such advancements in storage technology. At the Oxford Text Archive, for instance, a facility provided by the Oxford University Computing Services, over a gigabyte of disk space is needed to store over 1300 titles in its catalogue (as of 30 July 1993).

For a researcher of British or American literature, the databases of the aforementioned computing centers may be able to assist by acquiring machine-readable copies of the works he is studying. Sometimes upon request, these centers may provide free electronic copies of literary works to researchers via electronic mail, some of which, due to copyright restrictions, require permission from the original sources. A number of institutions, however, provide electronic texts mostly for public domain which can thus be downloaded from the Internet entirely free-of-charge; to mention a few, Project Gutenberg (e-mail: hart@vmd.cso.uiuc.edu), Cardinalis Etext Press (e-mail: keller@ra1.msstate.edu), and Internet Wiretap (e-mail: dell@wiretap.spies.com). Researchers can also make use of the search engines (e.g. www.yahoo.com, www.altavista.com, and www.lycos.com) to find the latest sites, using "e-text" as keyword, which offer electronic versions of literary texts free-of-charge. (One of the best ones is www.books.com.)

As of now, ten of the universities in the Philippines are connected to the Internet through a node of Phnet (www.pfi.net). Researchers who are not affiliated with any of these institutions can have Internet access through a number of Internet Service Providers (Philippines Online, Virtual Asia, Mosaic Communications, Epic Net, Globe Philippines, to name but a few). In Cagayan de Oro City, for example, a city located 505 nautical miles away from the capital, users can easily

register themselves with Oronet (www.oronet.com.ph), and access to this site will enable them to surf the Net, going in and out of countries around the globe without the usual visa requirements.

Of course, the more traditional researchers may still use snail mail to correspond with the above-mentioned computing centers; however, they would have to pay a fee for the diskette(s) as well as packaging and postage. Researchers who have CD ROM drives also have the option to acquire commercially available CD ROMs. Each with the capacity of about 650 high-density (HD) diskettes, a number of CD ROMs contain machine-readable versions of popular literary works.

In all probability, Philippine literary works may be among the last to be stored on computer disks in the world's computing centers; thus, a literary-statistical study on Edilberto K. Tiempo's use of color, for instance, would necessitate the creation of machine-readable versions of his short stories. Such a tedious process would undoubtedly dissuade many potential investigators. Since typing the text would consume much valuable research time (researchers need to spend their time on the more intellectually demanding matters), a scanner and computer software called OCR (optical character recognition) are desirable additions to a researcher's computing equipment. A scanner scans a page of a book to obtain an image of the page; the OCR translates the image into a string of letters and numbers, thus eliminating the need to type the text. Now, however, computer centers in Philippine universities, having evaluated priorities, do not have the best scanners for this purpose. In contrast, a number of business institutions in the country, particularly those engaged in desktop publishing, have added scanners to their inventory of computing paraphernalia. Some universities have acquired the so-called handheld scanners; designed to capture images, these scanners are not suitable for character recognition. Researchers in literary computing should opt for flatbed scanners, particularly those that have been demonstrated to work well with sophisticated OCR software. A currently popular OCR is OmniPage Pro which, using advances in neural network technology, can be trained to make smart decisions on indecipherable letters.

As an illustration of what suitable devices are capable of, the scanner and the OCR program available to the present authors can read a standard A4-size clear-copy page in 15 seconds. A book of normal size can therefore be scanned in less than a day.

Of course, the scanned text is not completely free from errors, especially when the pages of the book are old and difficult even for

an earthling to decipher. Such problems abound in the Philippines, where it is the custom of many publishers to use newsprint. Likewise, it is the custom of most copy machine owners to be sparing with ink and to use poor quality bond paper. All combine to reduce the chances of accurate scanning. In very extreme cases (the first author has encountered a few of them), hiring typists to key the text directly into the computer may be the most efficient approach. In such cases, researchers can take advantage of the cheap but relatively skilled Filipino labor.

It cannot be overemphasized that the results of a researcher's study depend on the accuracy of the inputted text. It is thus essential that texts generated by a scanner and an OCR be spell-checked, as should texts which have been typed in. A significant number of spelling mistakes would warn the researcher to proofread diligently.

Computer users will be aware that making electronic copies is very much more convenient (one does not have to flip pages), faster (it takes a little over a minute with older machines) and more compact (three regular-size books can be stored in one high-density diskette) than making hard (i.e., paper) copies of the same texts. However, when obtaining copies from other people, one has to be aware of the potential problems not only of incompatibility but also of computer viruses. (A program written by some prankster who delights in the suffering of others, a computer virus can be designed to destroy all of one's data. For instance, the virus may reformat the hard disk when the computer's clock exceeds a specified time on some date in the future; without warning, everything stored on the hard disk is then deleted.) A wise precaution is to run the latest version of an anti-virus program before using a diskette obtained from another computer user. In a country like the Philippines where pirating of software is said to be rampant, one can never be sure that even one's own hard drive has not been infected.

A useful tool is the spell-check feature included in most word processing programs. Some programs give the users the option to choose the variety of English with which to compare the typed-in text; for instance, the word 'organise' would be singled out as a misspelling in a program which uses American English, but not in a program whose standard is British English. Unfortunately, there is no program to validate Philippine English. Hired data entry professionals should be aware that although the word *aparador*, for example, is a misspelling in both varieties of English, it is legitimate in Philippine English. One can find instances of that word in the writings

511

of, to name but a few, N. V. M. Gonzalez, F. Sionil José, and Bienvenido N. Santos.[7] When a correctly spelled word is pointed out as a misspelling, word processors have the facility to add the word to the user's own personal dictionary; in so doing, the spell-checker treats the word as legitimate the next time it encounters that word.[8]

Nevertheless, researchers should not rely on spell-checkers alone: if the OCR software were to read 'comer' as 'comer', neither a spell-checker nor a grammar-checker (programs that check grammar have improved recently and have become cheaper too) would be able to spot the error. Furthermore, at present there are no spell-checkers for any of the Philippine languages, let alone Filipino or the so-called 'Taglish'. Although it is not an insurmountable task to write a computer program that checks for spelling errors in Filipino, creating the database of words with which to compare every typed word for possible misspelling is tantamount to keying in an entire Filipino dictionary. Thus, for investigations in Philippine vernacular literature, a researcher has no aids for proofreading the scanned material. He can of course add every vernacular word to his personal dictionary, but sooner or later this dictionary, having reached its maximum size, will start rejecting new entries.

Scanners, OCRs, spell-checkers, and proofreaders relieve the drudgery only of the preliminary stage of the investigation. Once all the text has been entered, special programs are needed to perform whatever textual analysis has been decided. Most of these programs are not readily available in the local computer shop, as they cater for rather esoteric needs. Again, a researcher would have to seek advice from one of the aforementioned literary and linguistic computing centers. These centers can sometimes supply programs free of charge. The Oxford University Computing Service, in particular, publishes a newsletter called *Computers and Texts* which not only provides up-to-date information regarding available software but also summarizes the research activities of other professionals in the field.[9] Unfortunately, because highly specialized computer software packages, such as many for literary and linguistic analyses, are sold only to an exclusive clientele, development costs cannot be amortized over a large number of sales. Consequently, potential buyers should not be surprised to find that the price of a single software item can sometimes exceed the cost of one IBM compatible personal computer assembled locally.

Because the acquisition of these text-processing programs may be beyond the budgets of many educational institutions in the Philippines, a more viable option might be to hire local programmers to

develop the software. Thus, universities which offer degrees in computer science or courses in programming could involve their students in such projects, possibly as part of what Philippine universities refer to as 'practicum'. This would be one of those rare opportunities for students in the humanities and the technical departments to get together. Universities which do not have students in computing could take advantage of the plentiful supply of skillful computer programmers in the country (the numbers, in fact, suggest that they form a significant labor-export). In contrast, in countries like Japan where labor is extremely expensive, hiring programmers for such a purpose would be considered ridiculous. Hence, researchers in the Philippines could actually benefit from software designed and subsequently maintained for their own investigations.

A very serious drawback is the need to train "the more or less innumerate majority of literary scholars" (Burrows 1992, 188). The root of this problem may be traced to the preparation of literature students, which "generally omits scientific training; thus, most critics feel inadequate in the face of essays or books on statistical assumptions or computational technology" (Potter 1989, xvi).

The following conversation overheard by a friend emphasizes that the problem extends to the Philippines:[10]

> Parent: OK, since your NCEE math score is low, why not take up education?
> Child: No, I'm not interested.
> Parent: Then major in English.[11]

Unfortunately such thinking appears to be prevalent, with the result that the level of numbers among students of the arts is a very serious impediment to the development of statistical studies in the humanities.

With current educational trends emphasizing interdisciplinarity (particularly in Ateneo de Manila University which has offered for many years a bachelor's degree in interdisciplinary studies), it is to be hoped that technical people would increasingly want to encroach upon the arts, and vice-versa. The eventual fading away of the present image of the literary scholar as devoid of numerical know-how would be one of the favorable offshoots of the integration of disciplines, thus breaking what appears to be a vicious spiral. For, indeed, the number of literary scholars afflicted with the myopic view that successful education means primarily a gain in literacy but not

necessarily in numeracy (i.e., that any innumerate can be considered 'educated' as long as he is literate) has grown to epidemic proportions.[12] Literary scholars compound the problem by passing this bias on to their students, thereby making the literary field, albeit unwittingly, a favorite meeting place for individuals who score low in mathematics.

Scholars in science and engineering would enjoy a similar advantage from interdisciplinary approaches: no longer would they be accused by such thinkers as José Ortega y Gassett (1951, 85–86) as, regrettably, being "learned ignoramuses" who know nothing save their "own tiny portion of the universe." In a quarterly departmental colloquium, Binongo (1993b, 1) says:

> I will begin this lecture by stating my biased idea of mathematics: more than a mathematician's plaything or a technologist's gadget, mathematics to me is first and foremost a liberal art. This bias is something recent. When I was writing my master's thesis years ago, I was working on the construction of a domain in C^2, which was in itself interesting because it was a domain that failed to exist in one-dimensional complex analysis. Finishing my degree and starting to work in the 'outside world', I often met people who after the usual exchange of words of introduction asked about my educational background. When I told them that I studied mathematics, they would then talk about one mathematical problem they took up in college or read in some magazine which I, despite my graduate degree, knew nothing save its name and its existence. The recurrence of such situations made me realize how specialized I have become even in the field which I call my own. I have made myself a victim of the phenomenon which the modern Spanish philosopher José Ortega y Gassett calls "the barbarism of specialization." Sure, being specialized has its own advantages: one gains more expertise in one's area by specializing. I, for one, will not go to an otolaryngologist if I happen to break my leg. My Japanese mentor sent me a card last Christmas, reminding me that "life is short; one cannot do everything one wants to do." Very true. But without having to sound like Ortega nor one of his disciples, specialization can sometimes be tragic, for it provides eyes that see only the foot of a mountain, not the mountain itself. If it was late autumn and the mountain was Mt. Fuji, one would have missed a most wondrous sight.

Matters deteriorate when literary scholars, aware that they have an aptitude towards figures, leave literary scholarship for good to graze on the greener pastures of the technical fields. Such a case is that of a literary professional (a product of Ateneo de Manila Univer-

sity) who, having supervised graduate research at Xavier University for several years, left the Department of English to head the Department of Accountancy, after she passed the grueling C.P.A. board examination. Indeed, the Department of English of that university (which used to be one of Mindanao's strongest) has experienced a relatively rapid turnover of good faculty.

In the Philippines one thus gets the strong impression that research in quantitative stylistics is confronted with many problems. However, there is no reason to despair. The field of computational stylistics is still at an early stage of development, and, as such, practitioners all over the world face similar problems. In the U.S., for instance, Potter (1989) reports of "literature PhDs moving into computing centers or departments of Computer Science to gain recognition for their work":

> Since there is little payoff in attempting to do criticism when fellow critics do not understand it, most critics who originally sought computer means to achieve literary critical goals very frequently turn away from criticism (xvi).

Thus if Philippine universities can take appropriate measures, the country should be able to catch up with the state of the art during the beginning of the next century.

Bien entendu, a prerequisite would be to raise mathematical standards at all levels. A subject of study in its own right, mathematics is also the language of science and technology. Furthermore, mathematics develops ways of thinking which are of broader benefit. It may be no accident that non-socialist countries with a strong mathematical tradition enjoy technological superiority and thus economic affluence.

Of course, it cannot be assumed that all students have an aptitude for the discipline. That is, the mathematics courses that students in the humanities take as part of their general education should be tailored to their own needs: the courses are designed to be as relevant and comprehensible as possible, particularly to those who do not have a flair for the subject. The implicit goal of their brief mathematical training is their eventual development as 'citizens of the world' and 'life-long independent learners' in mathematics. Unfortunately, the biggest obstacle is almost always the unwelcoming attitude many students have towards the subject. Ingrained and prevalent, such an attitude can hardly be expected to metamorphose into something less negative at the completion of two three-unit

515

courses at the tertiary level, unless teachers give up their traditional role as simply teachers of mathematics and start operating on the principle that they are essentially 'salespeople' of the discipline (Binongo 1995b). These teachers realize that a very significant aspect of their job is to attempt to make the subject more appealing to their students.

Mathematics which is relevant to literature students would include statistics. A few universities in the country have begun to move towards this direction. At Ateneo de Manila, for example, statistics is a required course for students pursuing the M.A. degree in Literature (English).[13] Needless to say, universities do not require their literature majors to take statistics just because they are dictated by a current fad to make this course accessible to students of all concentrations. Thus, squeezing humanities students in a statistics class designed for social science or education majors will do more harm than good; for literature students to appreciate the relevance of statistics, what they need is a course specifically oriented towards literary and linguistic computing. Anthony Kenny's (1982) lucid book, *The Computation of Style: An Introduction to Statistics for Students of Literature and Humanities,* illustrates how such a course could be designed; this book, incidentally, not only has been found useful by many practitioners but also has become the textbook in elementary statistics for many humanities majors. Burrows (1987, 11), himself an Emeritus Professor of Literature, recommends it as "an invaluable guide for beginners [in statistics] like me." Another text is Butler's *Statistics in Linguistics* (1985b).

In addition to statistics, courses in the history of mathematical thought, possibly incorporating courses in the history or philosophy of science, could be made available to students of literature. Courses such as these would bridge what hitherto has been a wide gulf in the understanding of human knowledge, as seen by the eyes of a humanities scholar and as by a practitioner of a technical discipline. They would fill a gap recognized by Bailey (1969, 232–33) who had complained that "educational national curricula seldom foster the development of the individual who is both 'numerate' and 'literate', despite the evident similarity in these two aptitudes." Aiming to instill among students a sense of, and a fascination with, interdisciplinarity, excursions into these regions sometimes can lead to a deeper understanding of the place of their respective disciplines. From such a perspective, quite original results can ensue. As an

illustration, Binongo (1993a), having attempted an examination of Nick Joaquin's literary humor from a mathematical perspective, concludes:

> The foregoing interdisciplinary adventurousness merely aims to demonstrate how a mathematical perspective can assist the literary critic in identifying and understanding the nature and form of a writer's humor as well as in categorizing those patterns, devices, and techniques the writer commonly uses to bring forth his desired effect. In addition to unraveling Joaquin's proclivity to stick to a humor device to compose his story, a mathematical perspective also helps put together what to an ordinary observer may seem separate if not unrelated concepts. Most important of all, by looking at literary humor with the eye of a mathematician, the critic not only realizes Koestler's idea, that an entirely different structure can be built on foundations that are similar to those of another, but also sees that understanding the design of a Gothic cathedral enables him to better understand the design of a cathedral designed by Gaudi.

More recently, David P. Ellerman (1995, 1–2) wrote an entire book, *Intellectual Trespassing as a Way of Life: Essays in Philosophy, Economics, and Mathematics*, to illustrate how intellectual trespassing (or what Binongo had called "interdisciplinary adventurousness") can in a number of ways be an engine of discovery:

> Thus dramatic changes or revolutions in a field of science are often made by outsiders or "trespassers" who do not define their expertise as mastery of the old methods or by newcomers who are not yet beholden to the old ways. That is one of the principal reasons why intellectual trespassing between fields should be promoted as a way of thought.
>
> There are other reasons for intellectual trespassing. The Polish mathematician Stefan Banach described mathematics as the science of analogies between analogies. Analogical reasoning is one of the most powerful engines of discovery. Trespassing is often inspired by noticing a new analogy between two different fields. Often the analogies are easily spotted because the same mathematics (which captures the analogies in abstract form) is used or can be used in both fields. Sometimes only basic knowledge in two seemingly disparate fields is needed to discover happy collisions of concepts that betray some common underlying structures. These cross-fertilizations may lead to new developments in one or the other of the fields, or they may only provide some intellectual amusement at the unexpected connections.

Indeed, students in science and engineering can also benefit from courses in the humanities designed to widen their understanding of human experience.

The *status quo* in a number of Philippine universities, however, is that there is a 'core curriculum' which stipulates the general education courses all students must take in order to graduate. In Ateneo de Manila University, the core curriculum at present includes, among others, 16 units of philosophy, 15 units of theology, 15 units of English composition and literature, 15 units of natural science, 6 units of a foreign language, 12 units of history, 9 units of Filipino composition and literature, 3 units of economics, 3 units of psychology, 3 units of sociology and anthropology, 3 units of political science and 6 units of mathematics for non-science majors or at least 16 units for science majors. Such a number of units in the humanities which an engineering student, for instance, takes is inarguably more than enough for his needs. Nevertheless, to gain maximum benefit from his university education, such a student would be well advised to take these subjects as seriously as his major subjects. A profound understanding of what appears to be irrelevant may prove to be of great value later in his life. The words of the chemist Henry Bauer (1991, 101) are apposite:

> Historians and sociologists of science have noted many instances of discoveries made by mavericks or by people who moved from one specialty into another—people who were able to see things hidden from those steeped in the old consensus. When isolation is broken, fresh viewpoints can be remarkably illuminating.

The philosopher Ortega y Gassett (1951, 86–87) has long understood that novel ideas can sometimes be traced to minds not resistant to interdisciplinarity.

> Newton was able to found his system of physics without knowing much philosophy, but Einstein needed to saturate himself with Kant and Mach before he could reach his own keen synthesis. Kant and Mach—the names are mere symbols of the enormous mass of philosophic and psychological thought which has influenced Einstein—have served to *liberate* the mind of the latter and leave the way open for his innovation.

The chemist Bauer (1991, 142) echoes the same message by asserting that the progress of scientific enterprise should not be a product of its practitioners working in isolation:

> The insights of practicing scientists and engineers are needed as much as those of historians, philosophers, and others if the nature of science is to be comprehended and the comprehension fruitfully acted upon.

The same breadth of perception is true for the understanding of the humanities.

Illustrations on Philippine Literature in English

This section provides three illustrations of what and how research in quantitative stylistics can contribute to literary scholarship in the Philippines. While the first two may require literary researchers to seek the guidance of statisticians and computer programmers, the third can be implemented readily by anybody familiar with the word search facility of word processors. These illustrations are neither exhaustive nor representative of the myriad variety of applications. The reasons for their inclusion in this section are that they analyze works in Philippine literature, that they exemplify different applications (actually, the first two employ a single statistical technique but for different purposes) and that they did not require a large commitment of research time. Hopefully, these illustrations will convince the skeptical reader of the utility of quantitative stylistics in understanding Philippine literary phenomena.

A quantification of qualitative differences. Though not primarily oriented to computer-assisted criticism, an article published previously in this journal, "Incongruity, Mathematics, and Humor in *Joaquinesquerie*" (Binongo 1993a), attempts to introduce the approach. Graphing the number of running words *(tokens)* on the horizontal axis and the number of different words *(types)* on the vertical axis, the steepness of the vocabulary curve of the *Joaquinesquerie (JQ)* corpus is reported to be much lower than that of Nick Joaquin's previous collection, *Tropical Gothic (TG)*. This supports the observation that Joaquin in his most recent compilation deliberately and drastically changes his writing style to one which uses a much less extensive vocabulary. (This is clear from Figure 1 on p. 479 of Binongo 1993a.)

In a paper published in *Literary and Linguistic Computing*, "Joaquin's *Joaquinesquerie, Joaquinesquerie*'s Joaquin: Statistical Expressions of a Filipino Writer's Style," the immensity of the previously noted gulf between the vocabulary curves of *TG* and *JQ* is shown: the gap yawns wide enough to accommodate the vocabulary curves of the

other four writers, N. V. M. Gonzalez, F. Sionil José, Bienvenido N. Santos and Edilberto K. Tiempo (Binongo 1994). As for Honoré's R, an index used in Binongo (1995a 81–82) to measure vocabulary richness, the same result is reported (Binongo 1994, 276).

In *"Tropical Gothic Versus Joaquinesquerie*: Quantifying Their Qualitative Differences," the differences between the two collections of stories are shown to also include sentence complexity and vocabulary content (Binongo 1995a). In line with Leech and Short's (1981, 67) view that "a rough and convenient measure of sentence complexity is the average number of words per sentence" this paper uses sentence length to demonstrate quantitatively that Nick Joaquin's latest collection of stories does not have his typical hyper-complex Faulknerian constructions. "Word content differentiation analysis reveals an even stronger result: the *Joaquinesquerie* stories reflect a lexical repertoire incompatible with traditional Joaquin's" (89). Thus this paper concludes:

> our experiments provide strong, supplementary—though this time, quantitative-based and objective—evidence that the collections are in diametrically opposite positions insofar as writing style is concerned.

The stylometrist is then warned:

> a stylometric analyst who wishes to demonstrate quantitatively that the two collections are of a single authorship will indeed be confronted with a gargantuan task.

Despite such words of caution, however, a subsequent paper succeeds in showing that *TG* and *JQ* are products of the same pen.

> The results were announced in Binongo (1993b, 1993c) and [were] published in ... Binongo (1994). Essentially, a multivariable statistical technique called "principal components analysis" is employed to analyze Joaquin's usage of thirty-six highest frequency words: "the," "and," "to," "a," "of," etc. [When] his works are juxtaposed with those of Bienvenido Santos, N. V. M. Gonzalez, Edilberto K. Tiempo, and F. Sionil José, ... regardless of the writing style he adopts and regardless of the stage of his growth as a writer..., an analysis based on the thirty-six most common word-types reveals the essence of Joaquin's writing style, a clear counterexample to the oft-stated position that everything but change changes." (Binongo 1995a, 90)

The alert reader will observe that as far as this application is concerned, the value of quantitative stylistics seems to lie in spectacularly demonstrating the obvious. Why does one have to go through all such intricate mathematical procedures if only to prove what is self-evident? Two decades ago, linguist Rebecca Posner (1963, 112) answered this problem:

> For some scholars·claim statistical studies cannot reveal anything that is not intuitively obvious: if they did, the revelations would be, by definition, irrelevant to literary worth, which can be judged only by its emotional or intellectual impact on the reader. This surely cannot mean that we demand that all statistical findings must be directly expressible in non-mathematical language. Though few of us would claim that ordinary language can give exhaustive expression to our emotional experiences—and mathematics can sometimes express the otherwise ineffable. However, if statistical studies were merely used for imparting precision to intuitive truths, we could not reject them on those grounds. Much of scholarship has just that aim, and is, none the less, worthwhile.

Moreover, in order to establish its legitimacy, quantification, as a different approach to literary analysis, must first demonstrate the validity and reliability of its methods. To do so, works of undisputed authorship as well as works which have been painstakingly examined by literary critics have to be used. Once the issue of validity and reliability is resolved, such techniques are ready for application to related problems. The unwary researcher is however cautioned; generalization is not automatic in studies in quantitative stylistics.

The by-products of quantification can be original and illuminating to, but not incompatible with, the already established coherent understanding of a writer's works. These by-products are often invisible to the eyes of even the most attentive reader. In one of the studies mentioned above, for example, it is symbolic that the word "God" tops the list of words (see Table 3 of Binongo 1995a, 84-6) which differentiate *TG* from *JQ*. The former is composed of stories that manifest Joaquin's "recurrent predisposition to moral and religious issues" (Binongo 1995a, 88), while the latter consists of stories which only reflect "his conscious intent to refashion traditional tales" (Binongo 1993a, 482) such as Shakespeare's *A Midsummer Night's Dream* ("How Love Came to Juan Tamad"), Browning's "The Pied Piper of Hamelin" ("The Hamiling Mystery"), the celebrated fairy

tales, "Beauty and the Beast," (Johnny Tiñoso and the Proud Beauty"), "Cinderella" ("The Amazing History of Elang Uling"), "Snow White and the Seven Dwarfs" and "Sleeping Beauty" ("The Mystery Sleeper of Balite Drive"), and so on. What is an even more surprising result is that an eigenanalysis of predominantly colorless words attests that Nick Joaquin

> hardly changes the identity of the thirty-six word-types over the forty years of his mature literary career. He may change his writing style from a rambling Faulknerian to a clipped fairy-tale type, from hyper-complex to simple sentence structures, or from a rich vocabulary level to a late grade-schooler's vocabulary level, but his usage of the highest frequency words remains relatively stable, apparently unaffected by his growth as a writer. ...Whether a reflection of ... "a self-gener-ated authorial individuality" or "a *tabula rasa* upon which larger cul-tural forces inscribe themselves," this regularity constitutes a curious result not only to literary critics but also to psycho/sociolinguists (Binongo 1994, 277).

This relative constancy in writing style is that which stylometrists try to isolate. Michaelson, Morton and Hamilton-Smith (1978, 20) ex-plain as follows:

> stylometry is not based upon the conscious and studied literary skills, no matter how long these have been practiced, how highly they have been developed nor how habitual they have become, but on a lower level of consciousness and a deeper level of organization, a framework which is established in childhood and which continues unchanged unless it is dissected and altered by conscious study, or is affected by disease or damage to the brain.

Unfortunately, the task of identifying this supposedly unchanging "framework" is anything but straightforward. The circumspect stylometrist realizes the myriad of extraneous variables he has to deal with: in what experimental psychologists call a 'confounded' inves-tigation (McGuigan 1190, 58–61), such variables as an author's level of maturity, the time of writing—or what Wellek and Warren (1956) call the "extrinsic" factors—and genre can all combine to render the results unreliable and invalid.

In addition to trying to isolate the 'signatures' of writers in their compositions, practitioners also aim to depict quantitatively how writers change their style over time. The use of the same technique

of principal component analysis shows the influence of the time dimension in both Edilberto K. Tiempo's and N. V. M. Gonzalez's short fiction. Resulting from an eigenanalysis of sixty-eight highest frequency words in the short story collections of the aforementioned five leading Philippine fiction writers, the writing of both Tiempo and Gonzalez is clearly affected by time.[14] In Tiempo's case, his collections are arranged from right to left, with the most recent one in the left-most. In Gonzalez's, the most recent one is in the right-most. Tiempo's writing style in his first collection of stories, *A Stream at Dalton Pass* (1970), is in stark contrast to the style of his three more recent collections. Gonzalez's writing style, on the other hand, remains steady until 1963. It is his latest stories, included in the collections *Mindoro and Beyond* (1989) and *The Bread of Salt and Other Stories* (1993), which exhibit a tangible stylistic shift.[15]

What has been captured by the statistical analysis may have been that which leaves the impression, often indescribable in words, in the minds of the readers that such-and-such a writer changes his style over time. Further research using non-quantitative critical approaches is recommended to confirm or dismiss this finding. There is no doubt, however, as to the chronological ordering of the short story collections based exclusively on the writer's usage of the most common words in English.

Again, this analysis demonstrates that function words, such as "the," "and," "to," "a," "of," "he," "was," "in," "his," and "it," deemed by both literary critics and linguists alike as aesthetically and linguistically uninteresting, are not really devoid of value. "The really nitty-gritty of the language of fiction is by no means confined to 'lexical words' " (Burrows 1987, 4). Burrows's (1992, 193) concluding remarks are relevant:

> There is no reason to suppose that those readers would expect such a picture to emerge from a statistical analysis of the most colourless word-types in the language. The point that deserves a closing emphasis is that, in this method of analysis, the word-types are allowed to 'choose' themselves, to interrelate at their 'choice', and to show up *whatever* mutual patterning is most influential as an expression of resemblances and differences within a given set of texts. [The graphs are] not a set of discrete entry-points but a closely woven network of meaning of a kind which we have scarcely begun to understand and which no one could have hoped to examine systematically before the advent of computers in literary studies.

An unraveling of apparently subconscious fixations. Literary scholars who shun statistical techniques in their analysis can still benefit from machine-readable texts by generating concordances of certain words. Strictly speaking, studies that use concordances are, to use Bailey's (1969, 230) phraseology, "more properly called *enumerative* than *statistical*." Nevertheless, this enumerative technique can yield interesting results; for example, a concordance of the words "naked," "nude," "strip" and their lemmas (that is, "nakedness," "nudity," "strips," and so on) suggests that nudity may be one of Nick Joaquin's fixations:

> Seldom pointed out in previous criticisms of Joaquin, stripping and nakedness are recurrent in the writer's earlier collections. Making the computer scan for the word naked in *Tropical Gothic*, we get the following examples: "Amada naked and screaming in bed" ("Summer Solstice"); "he imagined her in the posture of the idol and he stripped her"; "there she still was ... her breasts and shoulders naked" ("The Woman Who Had Two Navels"); "upon the bowed, mute, shriveled old man squatting motionless and cross-legged there, stark naked and half-blind and burned black as coal"; "Naked on the naked isle, he had ... pondered upon himself" ("Doña Jeronima"); "Here at the ends of the earth, alone under the skies, he had been stripped naked to the bone...." ("The Legend of the Dying Wanton"); "He... was stripped of jacket and shoes, tie and socks, shirt and trousers, undershirt and shorts. He was naked..." ("The Order of Melkizedek") The best—better yet, worst—example is the story, "Candido's Apocalypse," which contains a total of 36 occurrences of the word "naked." To cite but a few: "he had glanced around and seen her standing there stark naked"; "That was as far as he even got because he had seen the father inside stark naked, and the nakedness in a state of excitement"; "he was looking at a naked man playing with himself"; "but suddenly the clothes were not there anymore and he was looking at the naked man and he saw that everything in this man..." The naked in Joaquin's fiction is recurrent enough to be considered one of his minor, if not major, preoccupations. (Binongo 1993a, 509)

The advantages of using a computer to generate this concordance are obvious. Not only does it help to generate the concordance more quickly, the computer does not miss any occurrence of the desired words and, provided the researcher has anticipated all variants, their lemmas. Research results in cognitive psychology on human memory are relevant: first, with respect to short-term memory, it has been found that most people can hold seven, plus or minus two, items of information (Miller 1956 in the U.S., Ebbinghaus 1885 in Germany,

Yu 1985 in China). Long-term memory is likewise limited; cognitive psychologists agree that people do not store all of what they experience; neither are experiences permanently imprinted on the brain. What is even more striking is that people's memories are distorted by prejudices of all kinds (Loftus and Palmer 1974). In view of the limitations of their own memories as well as certain extraneous concerns which obstruct accurate recall, literary specialists on Nick Joaquin's fiction cannot be expected to recollect all the occurrences even of rare words in the entire corpus. In Nick Joaquin's case, the corpus extends over a period of more than forty years. Really, to resort to psychology is unnecessary; one only needs to rely on one's own experience of not being able to recall what one wrote to a friend, or even which friends one has written to during the past three years, to reach similar conclusions. A computer is certainly a welcome aid to memory.

Andrew Bennett and Nicholas Royle in their text, *An Introduction to Literature, Criticism and Theory* (1995, 22), argue that

> Even if we were to go to a living author [like Nick Joaquin] and ask what he ... meant by a particular text, all we would get would be another *text* (his... answer), which would then, in turn, be open to interpretation. Just because it comes 'from a horse's mouth' doesn't mean that the horse is telling the truth, or indeed that what the horse has to say about the 'words on the page' is any more interesting or illuminating than what anyone else might have to say. We could also ponder the question of 'authorial intention' in the light of psychoanalysis. 'Conscious intention', in this respect, can always be considered as subject to the unconscious workings of the mind. With psychoanalysis, it is no longer possible simply to privilege consciousness as the sombre judge of what is intended. The jurisdiction of 'authorial intention' falters here: what is not meant can still (in another sense) be meant.

Indeed, Joaquin himself may not have intended such a frequent use of the word "naked." A sampling of 200,000 words per writer, however, shows that for the words "naked" and "nakedness" alone, Joaquin's rate gives a figure of 69 times, followed by Edilberto K. Tiempo at 19 times, F. Sionil José 15 times, Bienvenido N. Santos 14 times, and finally N. V. M. Gonzalez with a modest 4 times. Joaquin's rate is 363% that of the writer next in line to him and 1725% that of the writer at the bottom of the list.

What about words that mean the same thing as "naked?" A frequency dictionary of Philippine English built by the first author,

based on the stories written by the writers above reveals that of all
the words synonymous with "naked" ("nude," "unclothed," "un-
dressed," "unclad," "stripped"), "naked" is most frequent. N. V. M.
Gonzalez uses the words *strip* or *strips* more often than any of the
other writers, but an examination of the corpus reveals that these
words appear in an entirely different sense, namely, "strips of colored
paper," "the week's comic strips," "a strip of fence," and so on. Sionil
José, Tiempo, Santos, and Gonzalez cannot match Nick Joaquin's
attachment to nudity. An investigation of Joaquin's concept of na-
kedness may therefore be revealing, perhaps in the same way as
Andrew Crosland (1977, 280–81) found from his concordance entries
for the word "house"—a noun that occurs frequently in *The Great
Gatsby*—that F. Scott Fitzgerald's "attached a symbolic significance
to it."

Another study which may benefit from a concordance is Edilberto
K. Tiempo's predisposition to numbers in his writing. This charac-
teristic separates him from other Philippine writers. Apparently una-
ware of this aspect of his writing, Tiempo in his correspondence with
the first author can only explain it by reference to what Henry James
calls the "solidity of specification."

Concordance-based studies can also be conducted within the works
of one writer. These are then stylistic studies of *internal deviation*. A
study of the frequency of "and" in Joaquin's short fiction shows that
it occurs at an abnormally low rate in "The Order of Melkizedek"
(Binongo 1994, 273). Consequently, an in-depth statistical-linguistic
examination may reveal interesting results regarding the writer's lin-
guistic use of this conjunction. Such a study may further demonstrate
how a colorless word like "and" may be able to enhance apprecia-
tion of nuances in the narration of the story.

When the desired words are interspersed with others, they may
lose some of their conspicuousness, making it difficult for the reader
to notice what appears to be a writer's subtle attachment to certain
ideas or modes of expression. Just as the reader may have been lost
in the forest of words, a writer preoccupied with trying to develop
his character, or build his plot to convey his theme, cannot be ex-
pected to be aware of apparently trivial recurring features in his text.
Thus, as Traugott and Pratt have been quoted in the beginning of
this article, creative writing does not always reflect conscious choices.
Scholars interested in unraveling the subconscious propensities, if not
conscious obsessions, of writers may take advantage of concordances.

Statistical Deviance, Psychological Prominence
and Literary Relevance

Indeed, studies in quantitative stylistics serve to encapsulate lin-
guistic evidence of a perceived style in a writer's works.

> The more a critic wishes to substantiate what he says about style, the
> more he will need to point to the linguistic evidence of texts; and lin-
> guistic evidence, to be firm, must be couched in terms of numerical
> frequency. (Leech and Short 1981, 47)

As mentioned in the preceding section, quantitative stylistics also
serves to provide

> confirmation for the 'hunches' or insights we have about style[;] ... it
> may bring to light significant features of style which would otherwise
> have been overlooked, and so lead to further insights. (47)

John B. Smith (1989, 39) adds:

> in addition to offering confirming examples, the critic may indicate the
> pervasiveness of that feature or pattern; by offering a comprehensive
> description of the features considered for the particular focus of the
> study..., the critic may address the question of the adequacy of a par-
> ticular assertion with regard to any specific combination of features.
> Thus, the computer offers the critic additional verificational concepts
> through its ability to address the entire text synchronically.

Although Leech and Short admit that "style is a such a compli-
cated phenomenon that it would be impractical to demand hard evi-
dence for every observation made," they are of the opinion that "it
may be sufficient for many purposes just to enumerate textual ex-
amples of the feature under discussion." These are readily extracted
from concordances. Sometimes, however, practitioners cannot refrain
from couching the observations in numerical terms. In trying to set-
tle a disagreement between two critics regarding Joaquin's consist-
ency in the use of "kilometric sentences that pile word upon word
and image upon image in almost breathless succession" (Roseburg
1966, 143), Binongo (1995a, 70) expresses the view that resorting to
quantification can actually help

> avoid gray areas that can lead to uncompromising views ... The em-
> ployment of a quantitative approach has the additional advantage of
> giving us tools to determine—in less ambiguous terms—the level of
> significance of any perceived difference in style.

There is, however, a more important reason for the need to express observations in numerical frequencies. Stated in purely qualitative terms, a study purporting to analyze a particular writer's style may possibly be founded on fallacious reasoning. Behavioral scientist Fred N. Kerlinger (1986, 4) explains:

> Nonscientists test "hypotheses," too, but they test them in what may be called a selective fashion. They often "select" evidence simply because it is consistent with the hypotheses. Take the stereotype: Blacks are musical. If people believe this, they can easily "verify" the belief by noting that many blacks are musicians. Exceptions to this stereotype, the unmusical or tone-deaf black, for example, are not perceived. Sophisticated social scientists, knowing this "selection tendency" to be a common psychological phenomenon, carefully guard against their research against their own preconceptions and predilections and against selective support of hypotheses.

The term 'selective evidence' will strike a responsive chord in researchers interested in the study of literary style. In regard to this cautionary note, Binongo (1993a, 504–5) adds that

> What would need explaining is not the occurrence of rare events but their nonoccurrence. Put in a different way, it is very unlikely that no unlikely event will occur. Juan Tamad's version would probably go: "It is not extraordinary to be extraordinary. On the contrary, it would be extraordinary to be ordinary."

That is, practitioners have to be extremely careful about making sweeping generalizations and magnifying the significance of rare or extraordinary phenomena.

The comparison of the rate of occurrence of the idea of nakedness in Nick Joaquin's short fiction with that of other writers eliminates the possibility that the perception could have resulted from selective observation. What about other statistical results (that is, those not obtained from concordances) which suggest literary insights that hitherto have been unnoticed? An example is the observation made earlier regarding how Tiempo's first collection of stories *A Stream at Dalton Pass* differs from his more recent collections. Leech and Short are of the opinion that a statistical statement like this proves nothing. To explain their view, they introduce three notions of saliency:

1. *Statistical deviance*: the difference between the normal frequency of a stylistic feature and the frequency of a writer's usage,
2. *Psychological prominence*: the degree to which a feature is salient to the reader's mind, and

3. *Literary relevance*: similar to the Prague School notion of *foregrounding* or *deautomatization* of the linguistic code, deviation that is artistically motivated.

Leech and Short argue that all instances of literary relevance are instances of prominence and that all instances of prominence are in turn instances of statistical deviance. The converse does not hold, however. That is, instances of psychological saliency can be due to factors other than literary considerations. For example, "Swift's dislike of monosyllables and Dryden's avoidance of final prepositions are cases of a writer's preferences being guided by a general sense of linguistic property, of what is 'good English.' "

> The dividing line between foregrounding and unmotivated prominence
> must be drawn in principle: where it is drawn in practice depends on
> a coherent literary interpretation of style. (Leech and Short 1981, 50)

Neither do instances of statistical deviance always constitute instances of psychological prominence.

> It is unlikely that certain deviances (below a level of significance) do
> not reach the threshold of response, even for the most experienced,
> alert, and sensitive reader. (49)

The conclusion then is that while it is true that statistical "deviance can be used to suggest and support hypotheses about style", "nothing can be adduced from, or proved by, statistics alone" (51).

Thus, while the statistical deviation observed in Tiempo's style can suggest further investigation for possible psychological prominence and literary relevance, on its own, statistical deviation is not of much importance. Nevertheless, the published literature indicates that statistical deviations sometimes provide excellent suggestions for more traditional literary approaches to the subject.

In the analogy used to counter Van Peer's philosophical objection, statistical analysis was likened to temporarily removing a part of the body in a surgical operation. This allows a meticulous examination of the part; for the body to resume its normal functioning, however, the part needs to be put back to its original place after the examination. Similarly, a quantitative surgery of a writer's style can become meaningful only after an understanding of the writer's actual texts has been acquired and when the statistical results are stated and in-

terpreted with due consideration of their literary context. For this reason, researchers in quantitative stylistics do not see computers as fully automating their work; in fact, their work is increased.

Quantification is a supplement; it cannot, on its own, purport to replace traditional study.

Leech and Short (1981, 71) conclude:

> Together, these arguments may seem to leave very little foothold for quantitative methods in the study of literary style. But on the other side there still remains the basic fact that without quantitative confirmation, statements on style lack the support of concrete evidence. Thus, although we are wary of claiming too much for statistical analysis, we would regard it as an essential and important tool in stylistic description.

In effect,

> just as the study of style cannot entirely rely on quantitative data, neither can it ultimately do without them (70).

Bailey (1969, 232) summarizes the details of how statistics can supplement traditional approaches to the study of literary style:

> [Statistical] techniques require a degree of explicitness in assumptions and procedures that is not highly valued in a discipline in which subjective and intuitive judgments are prized. Yet despite these apparent objections, statistical stylistics ... deals with questions that are of particular interest to the literary critic: Who wrote the work? In what directions did this writer develop? What are the constraints imposed on the writer by his language? How does the selection of the mode of presentation influence the shape of his work? How well do critics agree in their judgment of a piece of writing? Certainly these are all significant questions of vital interest to the literary scholar, and statistical methods can help him provide answers to them.

Conclusion

Literary-statistical studies do not require that researchers be well-versed in computing. But because lack of expertise does not preclude the ability to make use of computers, universities are at a disadvantage if they do not have up-to-date technological resources. With price tags now within the reach of the consumer, the cost of appropriate

hardware is generally not prohibitive (a state-of-the-art locally-assembled IBM compatible with an internal modem and a flatbed scanner would not cost over P50,000). Thus universities which offer courses in computer programming can now afford to fill a room with the latest computing equipment. Similarly, the needs of practitioners of computational stylistics are no longer unrealistic. As to software, universities can take advantage of their own students for the development of the required programs.

Nor is it a prerequisite that researchers be statisticians. Although an interdisciplinary field like computational stylistics presupposes an interdisciplinary interest, few can claim expertise in all its areas. Practitioners therefore need not be reticent in consulting specialists in the areas where they lack knowledge. Indeed, in order to demonstrate the legitimacy of this young discipline, practitioners cannot afford to turn research in this area into "a comedy of errors" as a result of pure naïveté! Shrewd practitioners of any interdisciplinary field bear in mind the saying, "Little knowledge is a dangerous thing." Or as Martin Luther King expresses it, "Nothing in the world is more dangerous than sincere ignorance and conscientious stupidity."

Editors Lubomír Doležel and Richard W. Bailey (1969, vii) note that

> stylistics and the study of texts have become a crossroads for the interests of literary scholars and critics, linguists, psychologists, sociologists, and mathematicians. The mathematician finds in the study of texts a body of material to use for testing and refining distributional and discriminatory formulas. The sociologist is interested in the text as it reflects social forces and developments in the history of ideas. The psychologist investigates the text as a product of the highest and most distinctive capacity of the human mind. The linguist studies the text structure in order to discover patterns and variability both in language system and the uses of language. Finally, the literary scholar and critic finds in the text the focus of his professional interest in artistic language.

Still,

> the statistical investigation of texts and text styles can serve all of the specialists just mentioned.

Doležel and Bailey (1969, vii) do point out, however, that

> though a scholar may be highly competent in his own area, he is often very much an amateur in one or more of the disciplines that find their meeting ground in statistical stylistics.

The branch of quantitative stylistics called stylometry is particularly susceptible. Tinkering with statistics as if it were taken from a recipe book, Merriam's pronouncement in 1980 that Shakespeare wrote *Sir Thomas More* resulted in shock waves of protest from the world of academic scholarship and is particularly remiss because the mistake could have been avoided had he sought statistical advice.[16] As another illustration, having examined O'Brien and Darnell's (1982) stylometric method in *Authorship Puzzles in the History of Economics*, Stephen Stigler (1983, 548) concluded that the results "must be dismissed as quite likely nothing more than artifacts of the procedures employed." Stigler ended his review with the following timeless reminder: "Quantification can be a dangerous matter itself" (550).

Potter (1989, xvi) recognizes that

> literary computing does not ... replace New Criticism's emphasis on the text as the central focus of study; indeed, it permits the closest possible examination of textual surfaces. Literary computing so little disputes structuralism that it in fact discloses structures invisible to the unaided eye. ... All these services and supports should, on the face of things, be of interest to traditional criticism; they are not replacements or supplantings of other methods but aids to the more convincing implementation of those methods.

Indeed, as Binongo (1995a, 66) acknowledges,

> far from admitting to be a necessary tool in literary analysis, the use of a mathematical perspective is however only a supplementary aid, as a magnifying lens is to a person. With the lens, a critic is able to see a particular aspect of a work more quickly; without the lens, a person's vision is in no way impaired.

A computational approach to criticism thus offers two advantages: that of providing supplementary and objective evidence to a problem which has been dealt with using the traditional approach, and that of suggesting to scholars of the traditional approach an in-depth qualitative examination of the matter which gave rise to the unexpected statistical results. In both cases, a complementary, rather than a competing or conflicting, relationship exists between the two approaches.

In conclusion, although an interdisciplinary field makes enormous demands on its practitioners, it brings about a greater consciousness of the Socratic humility: that the desire to be educated presumes an

acknowledgment of one's ignorance, that education is, after all, an endless beginning (Binongo 1985). Recognizing their own limitations, practitioners (both of the world of letters and of science and technology) with a willingness to learn may, through a quantitative study of literature, be able to achieve a better understanding of their respective specialisms. Concluding his review of the use of statistical inference in *The Oxford Shakespeare*, M. W. A. Smith (1991, 78) articulates the direction for future studies:

> Without a wealth of tradition and experience to draw upon, numerous apparently obscure and arcane pitfalls await an intrepid Shakespearian who ventures into interdisciplinary regions by invoking well-known statistical procedures without first examining their appropriateness to the intended application: just as the naivety of non-professionals who intervene in literary criticism can exasperate the scholar by insisting on eccentric interpretations, the endeavours of literati on scientific excursions can equally be invalidated because they are unaware of the standards of rigour essential for success in such work. Close cooperation between the Shakespearian, statistician, and computer scientist is surely long overdue.

Without the full cooperative effort of the scholars of the relevant disciplines, progress in interdisciplinary studies like quantitative stylistics is likely to be slow.

New developments geared towards the proper understanding of science illustrate the necessity of an interdisciplinary approach.

> This realization, that more than philosophy is needed to account for the success of science, that the explanation cannot rest on purely intellectual ground, is one of the reasons why some philosophers and historians and sociologists and others have come increasingly together in the scholarly interdisciplinary study (STS [science, technology and society]) of the activity that is science. One can comprehend science only by taking into account its history and its institutions, its social aspects as well as its cognitive ones; and the metaphors of puzzle and filter result from the determination to merge sociological with philosophical viewpoints. (Bauer 1992, 51)

Likewise, there is now a need for scholars of both literary and technical persuasions to join hands and be interdependent, demolishing centuries-old barriers and borders. New doors would then be opened to welcome the dawn of a different way of expanding and deepening human knowledge.

Notes

1. Readers interested in embarking on this type of study should write the first author at 106 Inoue Green Heights, 3-12-29 Muromi, Sawara-ku, Fukuoka 814, Japan. Alternatively, facsimile (81)-92-846-5949 or electronic mail JNG.Binongo@ulst.ac.uk. This article was completed in June 1995; it is therefore not entirely unlikely that a part of the technical discussion may have been outmoded by some recent advancement in computing technology.

2. "For various approaches to, and definitions of, style," Leech and Short (p. 40) advise the reader to refer to N. E. Enkvist (1964, 1973) and S. Chatman (1971).

3. Greenbaum (1991) further reports: "We were aware that computerized corpora were being developed for printed English in India (the Kolhapur Corpus) and in Australia (the Macquarie Corpus) on the model of the selection of samples in the Brown and LOB corpora, but (like the United States) these countries lacked a corpus of spoken English." (p. 84)

4. Now a professor of English at Waseda University in Tokyo, Dr. McFarland is still active in Philippine corpus linguistics and is currently building a much bigger corpus of Filipino.

5. Refer to the last sentence of note 1.

6. As the prefixes *kilo, mega,* and *giga* indicate, in computing parlance a kilobyte (KB) is roughly a thousand bytes; a megabyte (MB) is approximately a thousand KB or a million bytes; a gigabyte (GB) is about a thousand MB or a billion bytes.

7. Interestingly, in a sampling of over two hundred thousand words per writer, the word does not show up in Nick Joaquin's or Edilberto K. Tiempo's short fiction. Of course, one cannot deduce from this finding that neither Joaquin nor Tiempo uses the word in actual writing, let alone speech.

8. Even the word *technicalization,* which enjoys an entry in *Webster's* dictionary, is classed as a misspelling by the first author's spell-checker.

9. Interested readers may write to: *Computers in Teaching Initiative, Center for Textual Studies and Office for Humanities Computation, Oxford University Computing Service, 13 Banbury Road, Oxford OX2 6NN, U.K.* The present director is (Dr) Marilyn Deegan. This center is also accessible by fax (44) 865-273221 or e-mail *ctitext@vax.ox.ac.uk.*

10. The authors would like to thank Benigno D. Tutor, Jr. for providing this example and for relinquishing anonymity.

11. In his book, *Studying Literary Theory,* Roger Webster (1990, 1) reports that when he was studying for a degree in 'English Language and Literature' in the early 70s, "there was considerable attention given to the history of the English language, but little attention was paid to any of the issues which informed or underpinned the ways in which students read and study literature." Although, as Webster admits, much has changed in the British curricula since the 70s, in the Philippines one is faced with a different predicament which has continued to prevail up to this time: an English major in this country is usually a student in literature, not a student in the English language. And 'literature' does not necessarily mean literature written in English. A perusal of theses that have been awarded the Master of Arts in English degree at Xavier University reveals that studies conducted on vernacular literature have been accepted by the Graduate School of that university. The only thing English about these theses is that they were written in English. (Some theses, in fact, failed to provide English translations of the 'vernacular works under study.) But neither is that a hard-and-fast

rule for this university; the thesis of a former chairperson of the department (who obviously obtained her degree from the same university) was written in Filipino. In other words, 'English' in this university is a misnomer. Because it is rather a common practice, albeit wrong, for Filipinos to equate 'English' with any type of 'literature', it would not be surprising to see a similar pattern in other universities.

12. In his best-selling book, *Innumeracy: Mathematical Illiteracy and Its Consequences,* John Allen Paulos (1988, 3) complains about society's intolerance of a person's propensity to commit mistakes in grammar but not of the more prevalent and more serious inability of many people "to deal comfortably with the fundamental notions of number and chance." He writes:

> Later that evening we were watching the news, and the TV weathercaster announced that there was 50 percent chance of rain for Saturday and a 50 percent chance for Sunday, and concluded that there was therefore a 100 percent chance of rain that weekend. The remark went right by the self-styled grammarian, and even after I explained the mistake to him, he wasn't nearly as indignant as he would have been had the weathercaster left a dangling participle. In fact, unlike other failings which are hidden, mathematical illiteracy is often flaunted: "I can't even balance my checkbook." "I'm a person, not a numbers person." Or "I always hated math." (4)

13. Lit 295: Statistics and Literary Research is one of the three required courses; the other two are Lit 201: Methods of Literary Research and Lit 202: Literary Theory.

14. These words are "the," "and," "to," "a," "of," "he," "was," "in," "I," "his," "it," "had," "you," "she," "that," "her," "for," "with," "on," "not," "but," "at," "as," "him," "said," "they," "were," "be," "from," "there," "have," "is," "this," "all," "my," "would," "when," "me," "what," "now," "too," "one," "we," "could," "no," "like," "them," "out," "so," "who," "about," "or," "did," "their," "then," "by," "if," "are," "an," "been," "do," "man," "time," "your," "only," "back," "where," and "more."

15. Both collections, *Mindoro and Beyond* and *The Bread of Salt,* contain stories taken from earlier collections. Thus, only the stories written between 1970-90 have been included in the analysis.

16. Smith (1992, 438) blames Merriam's failure to understand the nature of a statistical test called chi-square "for all the misleading world-wide publicity in 1980."

References

Bailey, Richard W. 1969. Statistics and style: A historical survey. In *Statistics and style,* edited by Lubomir Dolezel and Richard W. Bailey. New York: American Elsevier Publishing Company, Inc.

Bauer, Henry H. 1991. *Scientific literacy and the myth of the scientific method.* Urbana: University of Illinois Press.

Bennett, Andrew and Nicholas Royle. 1995. *An introduction to literature, criticism, and theory.* London: Prentice Hall.

Binongo, José Nilo G. 1985. A constant beginning. Speech delivered at International Christian University, Tokyo (June).

——. 1993a. Incongruity, mathematics and humor in *Joaquinesquerie*. *Philippine Studies* 41:477–511.

——. 1993b. The literary and the mathematical: Attempts at a Philippine convergence. Paper read at the Mathematics Department Summer Colloquium, Ateneo de Manila University (May).

——. 1993c. A statistical way of discriminating Nick Joaquin's voice in Philippine short fiction. Paper read at the 20th Annual Conference of the Mathematical Society of the Philippines. Ateneo de Manila University (May).

——. 1994. Joaquin's *Joaquinesquerie*, *Joaquinesquerie*'s Joaquin: A statistical expression of a Filipino writer's style. *Literary and Linguistic Computing* 9: 267-79.

——. 1995a. *Tropical Gothic* versus *Joaquinesquerie*: quantifying their qualitative differences. *Philippine Studies* 45: 66–92.

——. 1995b. Why I teach the way I do: A statement of my educational philosophy. *Teaching Mathematics and Its Applications* 14: 107-11.

—— and M. W. A. Smith. 1996. Stylometry. *Notes and Queries* 43.4:448–52.

Bloch, Bernard. 1953. Linguistic structure and linguistic analysis. In *Report of the fourth annual round table meeting of linguistics and language study*. Edited by A. A. Hill. Washington, D.C.: Georgetown University Press.

Bratley, P. and D. Ross, Jr. 1981. Syllabic spectra. *Association of Literary and Linguistic Computing Journal* 2: 41–50.

Burrows, John F. 1987. *Computation into criticism: A study of Jane Austen's novels and an experiment in method*. Oxford: Clarendon.

——. 1992. Computers and the study of literature. In *Computers and Written Texts*. Edited by Christopher Butler. Oxford: Basil Blackwell Ltd.

Butler, Christopher. 1985a. *Computers in linguistics*. Oxford: Basil Blackwell Ltd.

——. 1985b. *Statistics in linguistics*. Oxford: Basil Blackwell Ltd.

Chatman, S., ed. 1971. *Literary Style: A Symposium*. Oxford University Press.

Crosland, Andrew. 1977. The concordance as an aid in the historical study of style. *Style* 11: 274–83.

Dolezel, Lubomír and Richard W. Bailey, ed. 1969. *Statistics and style*. New York: American Elsevier Publishing Company, Inc.

Ebbinghaus, H. 1885. *Memory: A contribution to experimental psychology*. Translated by H. A. Roger and C. E. Bussenius. New York: Columbia University Press.

Ellerman, David J. 1995. *Intellectual trespassing as a way of life: Essays in philosophy, economics, and mathematics*. Maryland: Rowman & Littlefield Publishers.

Enkvist, N. E. 1964. On defining style. In *Linguistics and Style*. Edited by N. E. Enkvist, J. Spencer, and M. J. Gregory. Oxford: Oxford University Press.

——. 1973. *Linguistic stylistics*. The Hague and Paris, Mouton.

Fortier, P. A. 1989. Some statistics of themes in the French novel. *Computers and the Humanities* 23: 298–99.

Gonzalez, Andrew. 1989. Preface to *A frequency count of Filipino*, by Curtis D. McFarland. Manila: Linguistic Society of the Philippines.

Gonzalez, N. V. M. 1947. *Seven hills away*. Denver: Alan Swallow.

_____. 1954. *Children of the ash-covered loam and other stories*. Manila: Benipayo Press.

_____. 1963. *Look, stranger, this island now*. Manila: Benipayo Press.

_____. 1989. *Mindoro and beyond*. Quezon City: New Day Publishers.

_____. 1993. *The bread of salt and other stories*. Seattle: University of Washington Press.

Greenbaum, Sidney. 1991. The development of the International Corpus of English. In *English corpus linguistics*. Eds. Karin Aihmer and Bengt Altenberg. London: Longman.

Holmes, David I. 1994. Authorship attribution. *Computers and the Humanities* 28: 87–106.

Kenny, Anthony. 1982. *The Computation of style*. Oxford: Pergamon Press.

Kerlinger, Fred N. 1986. *Foundations of behavioral research*. 3rd ed. New York: Holt, Rinehart and Winston.

Leech, Geoffrey N. and Michael H. Short. 1981. *Style in fiction: A linguistic introduction to English fictional prose*. London: Longman.

Lodge, David. 1966. *Language of fiction*. Routledge & Kegan Paul.

Loftus, E. F. and J. C. Palmer. 1974. Reconstruction of automobile destruction: An example of interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior* 13: 585–89.

Logan, H. M. 1982. The computer and metrical scansion. *Association of Literary and Linguistic Computing Journal* 3: 9-14.

_____. 1985. Most by numbers judge a poet's song. *Computers and the Humanities* 19: 213–20.

Lutoslawski, Wincenty 1897. *The origin and growth of Plato's logic*. London: Longmas, Green and Co.

Mendelsohn, Richard L. and Lewis M. Schwartz. 1987. *Basic logic*. New Jersey: Prentice-Hall, Inc.

McFarland, Curtis D. 1989. *A frequency count of Filipino*. Manila: Linguistic Society of the Philippines.

McGuigan, F. J. 1990. *Experimental psychology: Methods of research*. 5th ed. New Jersey: Prentice Hall.

Michaelson, S., A. Q. Morton and N. Hamilton-Smith. 1978. To couple is to custom. 4th rev. ed. Internal report. Department of Computer Science, University of Edinburgh.

Miller, G. A. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63: 81-97.

O'Brien, D. P. and A. C. Darnell. 1982. *Authorship puzzles in the history of economics*. London: Macmillan.

Ortega y Gassett, José. 1951. *The revolt of the masses.* 2d ed. London: Unwin Books.

Paulos, John Allen. 1988. *Innumeracy: Mathematical illiteracy and its consequences.* New York: Vintage Books.

Posner, Rebecca. 1963. The use and abuse of stylistic statistics. *Archivum Linguistsicum* 15: 111–39.

Potter, Rosanne G., ed. 1989. *Literary computing and literary criticism.* Philadelphia: University of Pennsylvania Press.

Sabol, C. R. 1989. Reliable narration in *The good soldier.* In *Computing and literary criticism.* Edited by Rosanne G. Potter, 207–23. Philadelphia: University of Philadelphia Press.

Smith, John B. 1989. Computer criticism. In *Literary computing and literary criticism.* Edited by Rosanne G. Potter, 13–44. Philadelphia: University of Philadelphia Press.

Smith, M. W. A. 1991. Statistical inference in *A textual companion* to the Oxford Shakespeare. *Notes and Queries* 236: 73-78.

_____. 1992. Shakespeare, stylometry, and *Sir Thomas More. Studies in Philology* 59: 434-44.

Stigler, Stephen. 1983. Review of authorship puzzles in the history of economics: A statistical approach. *Journal of Economic History* 43: 547-50.

Tiempo, Edilberto K. 1982. *Finalities.* Manila: New Day.

_____. 1970. *A stream at Dalton Pass and other stories.* Manila: Bookmark.

_____. 1988. *Rainbow for Rima.* Quezon City: New Day Publishers.

_____. 1992. *Snake Twin and Other Stories.* Quezon City: New Day Publishers.

Traugott, Elizabeth Closs and Mary Louise Pratt. 1980. *Linguistics for students of literature.* San Diego: Harcourt Brace Jovanovich.

Wellek, René and Austin Warren. 1956. *Theory of literature.* London: Penguin Books.

Webster, Roger. 1990. *Studying literary theory: An introduction.* London: Arnold.

Yu, B., et. al. 1985. STM capacity for Chinese and English language materials. *Memory and Cognition* 13: 202-7.