**Tropical Gothic Versus Joaquinesquerie:
Quantifying Their Qualitative Differences**

Jose Nilo G. Binongo

Tropical Gothic versus *Joaquinesquerie:*
Quantifying Their Qualitative Differences

José Nilo G. Binongo

In a previous article published in this journal, "Incongruity, Mathematics, and Humor in *Joaquinesquerie*," Binongo (1993a) illustrates bringing in mathematics to help understand literary humor. As subject, Nick Joaquin's latest compilation of stories, *Joaquinesquerie: Myth a la Mod* is used. 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. The question, thus, becomes: Can mathematics be of better service to literary analysis?

Skimming through the articles published in the journals, *Literary and Linguistic Computing* (Oxford University Press) and *Computers and the Humanities* (Kluwer Academic Publishers), we realize the question has already been answered in the affirmative in a variety of ways. The publication of such books as Burrows's *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method* (1987) even strengthens the point. Literary-statistical investigations of this nature have given rise to the field called *quantitative stylistics.* In a comprehensive essay, "Computers and the Study of Literature," Burrows (1992) writes that the main applications of quantitative stylistics lie in the areas of literary criticism and stylometry.

In literary criticism, word indices and concordances have been found useful, particularly in thematic analysis. Ariew (1978), for instance, has shown periodic patterns of fire and water images in Breton's surrealist prose poem "Poisson Soluble." Even words devoid

of content—the so-called "functional" or "grammatical" words—have been found not to be "devoid of meaning" (Burrows 1987, 29). Using 75 highest frequency words which are predominantly functional words ("the," "and," "I," "of," etc.), Burrows's (1992) own study suggests that "the language of English fiction . . . continued its passage towards modern vernacularity" (p. 191) and that, interestingly, clear differences between male and female writers during the eighteenth century become obliterated by the twentieth century (p. 192).

The other area, stylometry, is that which quantifies literary style primarily to resolve cases of disputed authorship. An open problem, for example, that may benefit from a stylometric examination is the authorship of the anonymous Elizabethan plays, *King Leir* and *The Taming of a Shrew*, which may or may not be earlier versions of Shakespeare's *King Lear* and *The Taming of the Shrew*. In *Theory of Literature*, Wellek and Warren (1956) stress the importance of these investigations:

> One of the first tasks of [literary] scholarship is the assembly of its materials, the careful undoing of the effects of time, the examination as to authorship, authenticity, and date. . . . Often the importance of these operations is particularly great, since without them, critical analysis and historical understanding would be hopelessly handicapped. (p. 57)

These scholars acknowledge that

> Statistical methods, mainly as to the occurrence and frequency of certain words, have also been used for the establishment of a relative chronology of Plato's dialogues by Lewis Campbell and especially by Wincenty Lutoslawski, who calls his method 'stylometry.' (p. 65)

Citing another example, Wellek and Warren continue:

> G. Udny Yule, a statistician and actuary, has used very complex mathematical methods to study the vocabulary of writers like Thomas à Kempis in order to establish the common authorship of several manuscripts. (p. 67)

Quoted from a textbook that has been, for decades, the bible of literature majors in Philippine universities, the passages above indubitably show that stylometric methods are not anything new. Stylometrists have been and are still developing more robust methods of authorship attribution (one of the more recent ones employs

a pattern recognition technique inspired by neurological research known as neural computation; see Matthews and Merriam 1993). Warren and Wellek conclude:

> Stylistic methods, if patiently developed, can supply evidence which, though falling short of complete certainty, makes identification highly probable. (p. 67)

Indeed, the field of quantitative or computational stylistics has come a long way. The agent that is primarily responsible for the rapid development of research in the field is the computer, which has become in recent years as "user-friendly" and accessible as one's personal diary or notebook (thus the coining of the term "notebook" to refer to computers of greatest portability). Pioneering studies—such as *The Statistical Study of Literary Vocabulary* (by the abovementioned Yule 1944)—were done in a period of our history when the computer was practically impossible to tame and thus left to the then-unenviable care of technical specialists. Quite understandably, these literary-statistical studies did not enjoy the popularity they now have—particularly in the United Kingdom where scholarship in this area has been very strong. Researchers today are privileged with the present unsurpassed growth of large stores of machine-readable texts, many of which are easily accessible by electronic mail or available in CD ROMs. Reputable universities have begun to set up institutes that provide facilities for this kind of research; the better-known are Oxford University Computing Service (UK) which is the largest, the University of Toronto's Centre for Computing in the Humanities (Canada) which probably houses the most versatile and best equipped group, Princeton and Rutgers Universities' Center for Electronic Texts in the Humanities (USA) which presently is headed by the founder and former director of the Oxford University Computing Service, and University of Newcastle's Centre for Literary and Linguistic Computing (Australia) which is under the direction of the aforementioned Burrows.

In the Philippines, computers have undoubtedly made quite a significant impact in educational institutions. However, in the literature departments even of the top universities, computers are used merely as word processors. Computer-assisted literary-statistical studies remain an untrodden territory.

Boldly pioneering it may be, this article initiates such an approach. Again, Nick Joaquin's short fiction will be used as subject.

Why Nick Joaquin?

Of Filipino writers in English, Nick Joaquin is noted for a peculiar writing style, so peculiar that identifying his work from that of another is often easy. Written for a much younger audience, his latest collection of stories, *Joaquinesquerie: Myth a la Mod* (1983), however, veers radically from his linguistic idiosyncrasies. Binongo (1993a, 478) has noted this drastic change in style:

> those who may have been conditioned to Joaquin's so-called "lush" writing style (Furay 1953, 152) are bound to be surprised. Hardly [in *Joaquinesquerie*] do we see an instance of those notoriously "long, kilometric sentences that pile word upon word and image upon image in almost breathless succession." (Roseburg 1966, 143)
>
> How *Joaquinesquerie* differs from a previous collection like *Tropical Gothic* can actually be quantified. . . . [T]he rate at which Joaquin uses new words in his stories in *Tropical Gothic* is much higher than that in *Joaquinesquerie*. Vocabulary richness is sometimes used as an indicator of a writer's maturity. Because the stories in *Joaquinesquerie* are less vocabulary rich, however much more recent (the gap in terms of the time of writing is in decades), than those in *Tropical Gothic*, we come to understand that Joaquin's change in his manner of writing in *Joaquinesquerie* is deliberate.

Binongo, whose mere intent was to offer an explanation for the stylistic departure, presumed the existence of wide linguistic differences between *Joaquinesquerie* and its writer's earlier fictional works, without actually going through a painstaking enumeration and examination of the details. It is the task of the present article to lay out these details. More specifically, this article investigates whether a quantitative approach can substantiate the stylistic differences between—made obvious by a mere reading of—Nick Joaquin's *Tropical Gothic* (1972) (which constitutes a representative anthology of his short fiction) and *Joaquinesquerie* (which is his most recent compilation).

Why a Quantitative Approach?

It is a known story that before the end of the Middle Ages, a bunch of theologians performed an experiment that had been prohibited by the Catholic Church but granted permission by a French

king. Wanting to determine the exact weight of the human soul, the "scientists" of that time weighed a criminal before and after he was hanged. A nonisolated case that typifies some "academic" investigations, this "experiment" concluded that the criminal's soul weighed an ounce and a half.

Even Leibnitz—revered today as one of the two fathers of calculus (the other is Newton)—in his youth believed in the primacy of quantification. He is often quoted for having declared in his naïve days that

> [i]f controversies were to arise, there would no more need of disputation between two philosophers than between two accountants. For it would suffice to take their pencils in their hands, to sit down to their slates, and to say to each other (with a friend to witness, if they liked): Let us calculate.

Clearly, man's attempts at quantifying what may be unquantifiable can easily become an object of laughter. We are thus impelled to ask: Are there aspects of a literary work that are easily and *naturally* quantifiable? Guerin et. al. (1992) in their popular book, *A Handbook of Critical Approaches to Literature*, remind us that

> no single approach can exhaust the manifold interpretive possibilities of a worthwhile literary work; each approach has its own peculiar limitations. (p. 116)

What then do we gain from departing from the more traditional approaches and from using a quantitative approach to literary analysis?

A mere reading of the *Tropical Gothic* (hereafter *TG*) and *Joaquinesquerie* (hereafter *JQ*) stories reveals a wide, unmistakable gap in writing style between the two collections. We note, however, that two decades ago critics who reviewed Joaquin's short stories before they were compiled into the *TG* collection disagreed on whether Joaquin's use of his style was constant. Lacaba (1968, 381), for instance, maintained that in the later *TG* stories Joaquin changed radically his writing style. Galdon (1976, 457), on the other hand, insisted that "a deeper analysis of style would reveal more similarities than differences with the earlier stories." In order to avoid gray areas that can lead to uncompromising views like this one, we resort to quantification. The employment 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. Admittedly, however, the 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 (which critics again and again have pointed out constitutes one of Joaquin's peculiarities as a prose writer; see, for example, Oloroso 1967): sound, imagery, and especially the figurative which, as literary scholars like Van Peer (1989) remind us, resists any attempt at statistical counts and frequencies. But neither can a nonquantitative approach boast of being comprehensive; the aspects of Joaquin's style examined in this article are those that are best expressed in quantitative terms.

Sentence Length

Sentence length defined as the number of words in a sentence is both the most tangible and one of the most easily quantified aspects of Joaquin's style. Joaquin's style is characterized, in the words of Furay (1953, 151), by "occasional long sentences which heap together, in one breath as it were, running impressions side by side with continuing action." While recognizing a crippling weakness of this style, Furay does not ignore the advantages:

> [Joaquin] has committed himself to the free exercise of this style, it seems, as that which all the rest builds toward; the subordinate parts of each story lift to the moments, here and there, when the meaning of the varied dramatic pictures he has been setting side by side can be distilled, full voice, in a poetic passage of sheer accumulated pictorial realization. So too throughout, the minor effects which prepare for the major are produced judiciously selected and vividly etched detail after detail thrust in a controlled stream on the reader's imagination. It makes for intense interest and for sharp realization at the moment when all the loose threads are gathered; and it is a technique which, at best, delivers the most stirring effects possible to that form called the short story. (p. 152)

Furay's remarks are in reference to the above-quoted story, "May Day Eve," and, in general, to the stories published in Joaquin's first collection, *Prose and Poems* (1952), from which six *TG* stories have been taken. Rephrasing Furay, Roseburg (1966) elaborates:

> [Joaquin's style] works best with the exotic, which is both its strength as well as its weakness because such a style demands that situations be unusual and that climaxes be violent, thereby forcing the action to fit the style and inevitably alienate the story from reality. (p. 143)

In response to Lacaba's (1968) observation that in two *TG* stories—
"Candido's Apocalypse" and "The Order of Melkizedek"—which
were written some thirteen years after the earlier *TG* stories, Joaquin
"changed radically his style," Galdon (1976) writes:

> Once the reader gets beyond the mere words of "Candido's Apoca-
> lypse" or "The Order of Melkizedek," the sentence patterns are still
> what Furay called "lush" and what I have called "baroque," or what
> the present author calls "Tropical Gothic." In that respect I see no sig-
> nificant difference between the two groups of stories. We are justified,
> then, in accepting the two groups of stories as representative of the
> same Early Joaquin, and *Tropical Gothic* as a representative anthology
> of Joaquin's short fiction of the period. (p. 457)

A mere reading of the stories would seem to indicate that Galdon is
correct. Here is a love scene taken from "The Order of Melkizedek,"
which, despite the preponderance of mod dialog, demonstrates that
Joaquin is not yet willing to break his habit:

> Kneeling to the pile, he thrust a kindled rag into it, then settled back
> on the grass beside her and took her in his arms. The fire grew slowly,
> the heart of flame burning buried by itself, yielding a coiled smoke as
> it slowly steadily dwindled, almost on the wane, only a smoldering,
> before it began to spread, gleaming into a line here, leaping into a burst
> of sparks there, the travelling tongues that swayed toward each other
> seeming to yearn for each other and for that heart of heat now swell-
> ing to include them too, now glowing into a furnace round which the
> darting sparks, shattering, multiplied the thicker smoke pushing in
> between brightening as, let loose, it uncoiled, unrolled, arose, rippling
> now from one blaze, one bush on fire, one total incandescence that
> rose, rushed, rustled, roared, raged, until, at last, finally fluently flow-
> ering, oh, oh, the burst body of it broke free, in a fountain of fire
> springing up to the skies; and they clung fast to each other, shudder-
> ing to the ascent.

While there can be disagreement over Joaquin's consistency in
sentence construction in the *TG* stories, there is no denying the sud-
den loss of traditional writing style in *JQ*. Reading the *JQ* collection,
one gets the impression that not only does Joaquin break in *JQ* the
stylistic habits he is noted for but that he entirely reverses his style,
taking it to the other end of the spectrum. The Filipino writer who wrote
stories with extremely complex sentence patterns now writes prose in
a manner much simpler than the average Filipino short fiction writer.[1]

Here is the beginning portion of a *JQ* story, "Lilit Bulilit and the Babe-in-the-Womb":

> That old Lilit! Always trying to look big!
> She was no bigger than a cat. (And she looked like a cat, too.) But how she liked to feel tall!
> She wore heels as high as herself. She wore a hair-do as tall as herself. She wore skirts or pants that trailed to the ground. She wore collars that came up to her ears. And the colors she loved to wear! Bright red, hot yellow, rich violet, glossy black. Never blue or white or pastel.
> And as thought that beehive hair-do of hers wasn't tall enough, she wore a pussy willow stuck on top of it. Imagine!

When juxtaposed with previous *TG* excerpts, a passage like this only leads to the conclusion that in terms of writing style the *TG* and *JQ* stories are more different than alike.

Of course quoting paragraphs here and there from the *JQ* stories cannot constitute an argument for Joaquin's use of much shorter sentences in *JQ*: in the *TG* collection too, he uses dialogue, and dialogues do not constitute one-sentence paragraphs. Table 1 lists the sentence length statistics of *TG* and *JQ*; the information here should provide an overall picture of the situation.

The distribution of sentence length, as we know it, is positively skewed (or skewed to the right), with shorter sentences tending to recur more frequently than the longer ones. The sentence distribution of the *TG* and *JQ* corpora does not break this law (see Figure 1). Table 1 clearly shows that the mode of Joaquin's sentence length is always smaller than his median which in turn is smaller than his mean. The important thing to look at in Table 1, however, is not how Joaquin subscribes to some natural law but rather how his sentence length statistics in *TG* and *JQ* differ.

Kilometric sentences used only in TG. Our first remark has been noted previously: not a single occurrence of those "notoriously long, kilometric sentences" can be found in *JQ*. Writing the *JQ* stories then, Joaquin breaks decades of habit overnight:

> **Bresnahan:** [When you were writing your children's stories], did you find that you had to keep revising?
> **Joaquin:** I knew right away what language I would have to use, and I wasn't writing down to children. If I tended to write long sentences, then I started writing shorter ones. (Bresnahan 1990, 63)

73

More specifically, the longest sentence in *JQ* (65 words) is an eighth of the greatest maximum sentence length (530 words) in *TG* and is half the lowest maximum sentence length (128 words). When a difference of this order arises and the level of variability is kept to a minimum, no recourse to statistical tests of significance is necessary.

Higher sentence length means in TG. Most *TG* stories have a higher sentence length mean than most *JQ* stories. Combining the *TG* stories into one corpus and the *JQ* stories into another, we see that the average sentence length mean of the resulting *TG* corpus is higher than that of *JQ*, a difference of about 5 words per sentence.

More diversity in sentence length in TG. Another difference between *TG* and *JQ* is the degree of variability in both intrastory and interstory sentence length. The standard deviations of the *TG* stories are all higher than those of the *JQ* stories. This means that each *TG* story has a more variable sentence length than each *JQ* story. Because the standard deviation is sensitive to extreme values, we can say that this difference in variability can be attributed in large part to our first observation: by adding unusually long sentences in every *TG* story, Joaquin produces more variability in sentence length in each *TG* story.
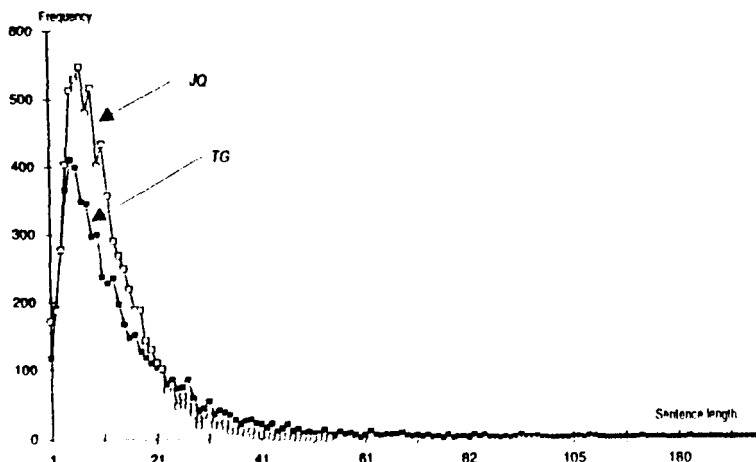

Fig 1 Sentence Length Distribution of the *Tropical Gothic* and *Joaquinesquerie* corpora

The differences among the means, modes, and medians of the *TG* stories as compared to those of the *JQ* stories also reveal interstory variability—that the *JQ* stories are much more like each other than the *TG* stories are to each other. It may be suggested that this difference is due to the fact that the *TG* collection contains writing that spans two decades—from 1946 to 1966—while the *JQ* stories were written only in a year. But this observation also admits of the fact that Joaquin does not really have a monolithic sentence length record even in his early stories. In other words, the only thing consistent about the nature of Joaquin's sentence length in *TG* is that in all the *TG* stories Joaquin once in a while inserts kilometric sentences; however, statistical measures of central tendency reveal more variety than consistency in the writer's handling of sentence length. Contrary to what readers usually remember of him, Joaquin's kilometric sentences do not constitute his mean, median, or mode in his individual stories; rather, they merely reflect his constant but infrequent flair for rambling Faulknerian constructions.

What could have happened had Joaquin avoided the use of kilometric sentences in *TG*? This is exactly what happened to the *JQ* collection. Invoking our first observation again, we posit that the absence of long sentences in *JQ* has made it easier for Joaquin to maintain a stable sentence length pattern in the *JQ*.

The story, "May Day Eve," is worth singling out: it is a *TG* story—its longest sentence being over 250 words—but its sentence length mode is the smallest of Joaquin's stories, both *TG* and *JQ* combined. It also happens to be a very short story, in fact shorter than any of the *JQ* stories.

To resolve the disagreement between Lacaba and Galdon, we note that there appears to be some truth to Lacaba's observation that Joaquin changes his style in "Candido's Apocalypse" and "The Order of Melkizedek." The sentence length modes of these stories are 5 words per sentence, which is a very low figure for a *TG* story. Lacaba, however, misses the fact that Joaquin's short sentences appear most frequently not for the first time in these two *TG* stories, but in such early stories as "Summer Solstice" and "May Day Eve" with a frequency of 4 and 5 words per sentence, respectively. Thus, if Lacaba was indeed referring to sentence length mode in his remark, we cannot but accept it with reserve. Galdon, on the other hand, is correct in the sense that Joaquin still uses long sentences with their corresponding complex sentence patterns in the later *TG* stories. In other words, Lacaba may have referred to the frequency

Table 1   Sentence Length Statistics of Joaquin's Stories

Story	Length of Longest Sentence	Number of Sentences	Sentence Length Mean	Standard Deviation	Sentence Length Mode*	Sentence Length Median
Candido's Apocalypse (TG)	417	1257	17.10	24.46	5, 94	10
Dona Jeronima (TG)	513	390	26.63	35.41	10, 21 17, 21	19
The Legend of the Dying Wanton (TG)	530	174	31.81	56.42	12, 11	20
May Day Eve (TG)	261	257	14.89	30.00	5, 28	7
Summer Solstice (TG)	164	366	12.80	14.00	4, 35	8
Guardia de Honor (TG)	173	488	12.86	13.92	8, 40	9
The Mass of St Sylvestre (TG)	128	101	26.34	18.12	15, 8	23
The Woman Who Had Two Navels (TG)	190	909	20.16	20.76	6, 52	13
The Order of Melkizedek (TG)	151	2235	11.99	9.33	5,162	10
The Traveling Salesman and the Split-Woman (JQ)	43	509	10.09	6.59	6, 49	9
The Mystery Sleeper of Balite Drive (JQ)	45	333	12.78	7.12	9, 30	11
Going to Jerusalem (JQ)	47	470	10.41	6.48	5, 48	9
How Love Came to Juan Tamad (JQ)	42	520	10.51	5.99	9, 50	9
The Hamiling Mystery (JQ)	58	624	13.50	8.33	7, 44	12
Lilit Bulilit and the Babe-in-the-Womb (JQ)	63	542	10.87	7.76	7, 47	9
The Four Little Monkeys Who Went to Eden (JQ)	62	468	13.87	7.95	9, 36	12
Sarimanok Versus Ibong Adarna (JQ)	45	558	9.56	6.35	5, 52	8
The Adventures of Culas-Culasito (JQ)	65	403	11.89	8.32	7, 34	10
The Amazing History of Elang Uling (JQ)	47	386	12.61	8.92	5, 26	11
Gotita de Dragon (JQ)	59	441	11.93	7.55	6, 35	10
The Happiest Boy in the World (JQ)	41	562	9.78	6.79	5, 55	8
Johnny Tinoso and the Proud Beauty (JQ)	49	386	12.21	7.13	9, 32	11
Balikbayan (JQ)	42	593	10.46	7.01	8, 48	9
TG Corpus	530	6177	16.19	21.94	5, 410	11
JQ Corpus	65	7262	11.24	7.39	7, 547	10
Both TG and JQ	530	13439	13.51	16.02	6, 927	10

*i, j: i is the mode; j is the number of occurrences.

Table 2   Indices of Vocabulary Richness of Joaquin's Stories

Story	Tokens	Types	Ratio	Honore's R
Candido's Apocalypse (*TG*)	21504	3379	15.71%	988.49
Dona Jeronima (*TG*)	10386	2139	20.60%	986.36
The Legend of the Dying Wanton (*TG*)	5534	1563	28.24%	994.96
May Day Eve (*TG*)	3827	1015	26.52%	874.18
Summer Solstice (*TG*)	4686	1287	27.46%	1003.04
Guardia de Honor (*TG*)	6277	1379	21.97%	824.74
The Mass of St Sylvestre (*TG*)	2660	949	35.68%	1120.76
The Woman Who Had Two Navels (*TG*)	18321	3466	18.92%	1059.17
The Order of Melkizedek (*TG*)	26791	4352	16.24%	1007.35
The Traveling Salesman and the Split-Woman (*JQ*)	5137	1130	22.00%	794.15
The Mystery Sleeper of Balite Drive (*JQ*)	4257	1048	24.62%	839.58
Going to Jerusalem (*JQ*)	4893	1029	21.03%	787.67
How Love Came to Juan Tamad (*JQ*)	5469	1199	21.92%	836.15
The Hamiling Mystery (*JQ*)	8421	1800	21.38%	926.04
Lilit Bulilit and the Babe-in-the-Womb (*JQ*)	5896	1209	20.51%	770.03
The Four Little Monkeys Who Went to Eden (*JQ*)	6493	1447	22.29%	855.29
Sarimanok Versus Ibong Adarna (*JQ*)	5337	1193	22.35%	781.49
The Adventures of Culas-Culasito (*JQ*)	5496	1169	21.27%	828.05
The Amazing History of Elang Uling (*JQ*)	5261	1071	20.36%	776.86
Gotita de Dragon (*JQ*)	4870	1309	26.88%	983.09
The Happiest Boy in the World (*JQ*)	4790	1111	23.19%	816.14
Johnny Tinoso and the Proud Beauty (*JQ*)	4715	1198	25.41%	907.39
Balikbayan (*JQ*)	6208	1421	22.89%	905.85
Lechonito, the Holy Innocent (*JQ*)	4357	1156	26.53%	834.70
TG Corpus	99986	9796	9.80%	979.39
JQ Corpus	81600	6888	8.44%	886.11
Both *TG* and *JQ*	181586	12559	6.92%	941.40

of short sentences which is the mode—this observation is not without glaring exceptions; Galdon, on the contrary, may have been talking about the occurrence of long sentences. And we have seen that it is not the frequency of long or short sentences that distinguishes between Joaquin's early and later works in *TG*, not even between the *TG* and the *JQ* collections; it is simply his sporadic, if not temperamental, addition of extremely long sentences that holds the *TG* stories together, marking them off as non-*JQ*.

Vocabulary Richness

Statistics on sentence length do not reveal anything about the more important parameter of a writer's style—the word. In a doctoral thesis submitted to University of Cambridge, Tallentire (1972) suggests that stylistic investigations begin at the lexical level for the obvious reason that more data exist in this level.

Analysis at the lexical level has prompted Binongo (1993a, 478) to contend that the *TG* stories are vocabulary richer than the *JQ* stories. (Table 2) How does one measure the level of vocabulary richness of a literary work?

Type-Token Ratio

One simple index that is used widely, though too often misused and abused, is the *type-token ratio*. This is the ratio of the number of different words to total number of running words. By definition then, the higher the ratio, the higher the tendency for a writer to use different words in his text. It is for this reason that the type-token ratio is sometimes used to measure vocabulary richness.

What is meant by "different words"? If we count words like "be," "am," "is," "are," "was," "were," "been," "being," i.e., words that belong to the same lexeme, as separate words, we obtain the so-called *unlemmatized* type-token ratio. A *lemmatized* type-token ratio could have been resorted to, but in this article we count graphic words, i.e., we count words as they literally stand in the texts. This also means that homographs (e.g., "to" before a noun or a pronoun as a preposition and "to" before a verb as an infinitive) are not separated, nor are contracted forms (e.g., "we'll") expanded.

Baker (1988, 36) asserts that "an author's rate of generating new words is reasonably independent of the word length of his manuscript—once the text has passed certain limits." That two variables

are independent implies the absence of any significant correlation between them. (The converse is not true: the absence of a significant correlation may only indicate an absence of a linear relationship.) A correlation coefficient of +1 signifies a perfect match, -1 a perfect mismatch, and 0 a perfect absence of relationship. The correlation coefficient between text length and the type-token ratio[2] of Joaquin's stories is -0.68, and this figure albeit not perfect is significant, warranting a rejection of Baker's contention.[3] (The concept of statistical significance will be explained in the next paragraph.) Thus, because there is a strong dependence between type-token ratio and the number of tokens, a given ratio cannot stand by itself. For a given type-token ratio, one has to be conscious always of the text length used to calculate the ratio.

Insignificant difference in text length between TG and JQ. There is a difference of 5,670 words between the mean text lengths of *TG* and *JQ*. Is this difference significant? If we consider that 5,670 words per story is greater than the mean text length of *JQ*, which is 5,440 words, we may be led to believe that the difference in text length between the two collections is significant. But because the variances are too large, especially among the *TG* stories, we need to perform a statistical test. As a first step, let us hypothesize that *TG* and *JQ* come from two populations, the text length means of which do not differ considerably. We are interested in finding out whether or not we can find sufficient evidence in our data that is inconsistent with this hypothesis, the so-called "null hypothesis." If we do, we would have to reject the null hypothesis and accept "the alternative hypothesis" that the text length means of the populations represented by *TG* and *JQ* are not the same. Because the variances of the samples are too far apart from each other,[4] we use, in lieu of t, the t' statistic which is approximated by the Behrens-Fisher distribution. Using the proposed method of Cochran and Cox (1957), at $t' > 2.31$ the probability of wrongly rejecting the null hypothesis when it is true is less than 5%. Because this probability is small, we take the region $t' > 2.31$ to be the "region of rejection," i.e., the region where we can feel fairly confident about rejecting the null hypothesis. The t' value we get, however, is 1.85, and this falls outside the region of rejection. As such, our data provide insufficient evidence to reject the null hypothesis; i.e., we cannot say at the 5% confidence level we have set that the text lengths of the populations represented by the two collections differ.

Insignificant difference between the TG and JQ type-token ratios. The difference between the *TG* and *JQ* type-token ratios is only 23.48%-22.84% = 0.64%. This difference is insignificant even at a 20% level.[5] One may say that the nearly nil difference between the type-token ratios of the two collections attests to the fact that the writers of the two collections are the same person. In other words, the type-token ratio contains authorship information. This makes a wonderful result, but if we recall our earlier demonstration that the type-token ratio is not stable with regard to text length, we can never be very sure. That is, we cannot guarantee that the aforementioned absence of a significant relationship between text lengths of the two collections does not have a not-so-insignificant impact on the absence of a significant relationship between their type-token ratios. We do realize, however, that we have to reduce the noise generated by text length on any index of vocabulary richness.

Type-Token Rate

One natural way to circumvent the effect of text length on the type-token ratio is to keep the text length fixed when comparing type-token ratios of different texts. Moreover, instead of associating a text with a single ratio (as Baker does), we will pay more attention to how the ratio changes as we move from one point in the text to another. In other words, we will not be looking at the static type-token ratio per se but at the dynamic, for lack of better term, *type-token rate*.

In order to monitor changes in the type-token rate R of a text, it is best that we draw its vocabulary curve, plotting the number of tokens N in the horizontal axis and the number of types V in the vertical axis.[6] Figure 2 (as well as Figure 1 in Binongo 1993a, 479)—and almost all vocabulary curves, for that matter—shows that the relationship between N and V tends to obey, to import a term from economics, the law of diminishing returns: at first V increases as fast or even faster than N, but after a certain point, N can no longer cope with the increase in V, resulting in a lower and lower R.

Higher type-token rate in TG corpus; higher type-token rate in TG stories. As noted previously, the rate at which Joaquin generates new words is much faster in the *TG* than in the *JQ* corpus. In fact, *TG*'s superiority over *JQ* increases with the number of tokens, as evidenced by the increasing gap between their vocabulary curves (see Figure 1 in Binongo 1993a, 479). Figure 2 plots the vocabulary curves of the

nine *TG* stories and the fifteen *JQ* stories. The vocabulary curves of the *TG* stories are generally steeper than those of *JQ*.

It is, moreover, worth noting that looking at the type-token ratios of the *TG* and *JQ* corpora, we get a result that is incompatible with our earlier observation that longer texts tend to have lower type-token ratios. This only suggests that the level of vocabulary richness of the *TG* stories is high enough to offset the negative effect of the law of diminishing returns.

Honoré's R

An index of vocabulary richness that reduces the effect of text length is *Honoré's R*.[7] Holmes (1992) explains:

[Honoré's *R*] directly tests the propensity of an author to choose between the alternatives of employing a word used previously or employing a new word. In the extreme case when each word type in a text is used once only, $V_1 = V$, and *R* becomes infinite. When comparing texts, therefore, the higher the *R*-value is, the richer the vocabulary in the sense that a greater number of words appear infrequently. (p. 93)
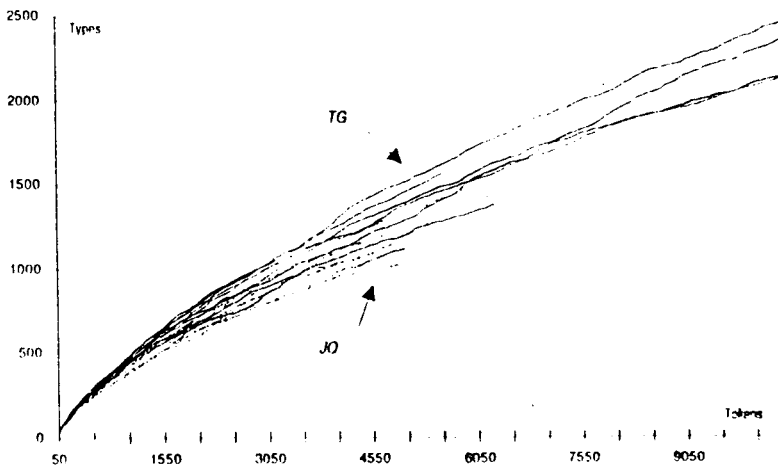

Fig 2   Vocabulary Curves of Joaquin's Stories

Moreover, R is "independent of text length N, a most important attribute" (Holmes 1991, 264).

Now that R is said to effectively neutralize the effect of text length, do we see a significant difference in Honoré's R between the two collections? The answer is affirmative; in fact the difference of 984.34 - 842.83 = 141.50 is significant.[8] Thus, the level of vocabulary richness of the *TG* stories as measured by Honoré's R is significantly higher than that of *JQ*.

Word Content

As we go deeper in our analysis of the differences between Joaquin's two collections, we cannot help asking exactly what words in *TG* are used in a significantly different frequency from that in *JQ*, and vice-versa. This requires that we go beyond mere indices on vocabulary richness and delve into the actual words.

Our first step is to get the raw frequencies of each word for each story. To standardize these frequencies, we divide each of these frequencies by the total number of words, expressing the new frequencies as percentages. We then compare the means of the *TG* and *JQ* stories for each word that Joaquin uses. The difference in frequency of usage of words such as "do," "rare," "sunshine," "was," "always"—the corresponding t-scores of which are 0, 0, 0, 0.001, and -0.001—between the two collections is not significant. We are more interested, however, in those *TG* words that are used differently in *JQ*, and vice-versa. Table 3 lists the words for which the t-scores are significant at 2.5% level.[9]

Based on this list, we can make a few observations about the differences between *TG* and *JQ* in regard to word content. Below we list some, after which we try to frame the separate observations into one picture.

Preoccupation with religion, morality, and original sin in TG. A mere glance at the list shows most of the *TG* words that significantly differ in usage from *JQ* are related to religion, morality, and sin. To name a few, we have "God"—which enjoys the most positive t-score, "sacred," "grace," "forgive," "mercy," "soul," "life," "intoned," "imploring," "enchanted," "saints," "priest," "mystic," "Christian," "Christ," "faithful," "religion," "processions," "peace," "honest," "piety," "decent," "respected," "properly," "conscience," "patience," "virtue," "dignity," "manners," "processions," "hell," "flesh," "naked-

ness," "stripped," "sins," "gods," "damned," "vanity." The prevalence of these words may lead us to believe that the *TG* stories are religious. Critics like Casper (1966, 137–39) disagree: Joaquin's *TG* stories "are not Christian so much as fundamentally pagan is their devotion to superstition. . . . Joaquin's [*TG*] stories chiefly derive not from within the Mystical Body . . . but within the shifting lore of folk not fully converted." It cannot be argued, however, that this dimension of the *TG* stories, whether one calls it Christian or pagan, is significantly reduced in the *JQ* collection. As noted in an earlier article, the mingling of the Christian and the pagan is still present in the *JQ* stories—in fact, Binongo (1993a) has posited that Joaquin uses this as one of his humor devices in the *JQ* collection—but incontestably the religious words in *TG* no longer enjoy their conventional popularity in *JQ*.

Words of degree and intensity—strength vs. weakness, passion vs. indifference in TG. This is the second group of words that differentiate *TG* from *JQ*. Words like "entire," "pleasure," "desire," "desired," "vitality," "vigorous," "passion," "intimate," "close," "gay," "fierce," "splendid," "aloud," "feeling," "power," "fires," "lovers," "thrust," "action," "pursuit," "amazement," "awed," "infinite," "gaily," "gravely," "swiftly," "instantly," "scurrying," "flowing," "surging," "leaping," "panting," "blazed," "ripped," "increased," "cluttered," vs. "mere," "perhaps," "weary," "moveless," "vague," "sober," "frail," "slightly," "merely," "coldly," "vaguely," "equally," and "limply" abound in the *TG* collection.

Macabre words in TG. Examples are "blood," "bloody," "bleeding," "rot," "bodies," "savage," "torture," "throat," "dreadful," "bones," "warrior," "brutal," "cruel." These words make one recall Roseburg (1966) who has asserted that

> [Joaquin's] realism is neither polite nor gentle but brutal and raw-boned,
> and at times shocking and sordid . . . [It is a] realism that makes the
> flesh creep, the blood course violently, and the pulse quicken. (p. 141)

Words of perception, darkness and light in TG. Another group of words are the following: "eyes," "face," "hair," "nape," "apparel," "clothing," "gesture," "audience," "lamps," "windows," "walls," "balconies," "lightning," "witness," "behold," "perceived," "shrouded," "smiling," "glittering," "white," "visible," "vividly," "dimly," "mood."

Words of calmness vs. disturbance in TG. Examples are "calm," "relaxed," "quietly," "tumult", "baffled," "worrying," "crisis," "torn,"

Table 3   Most Significant Words in *Tropical Gothic and Joaquinesquerie*

TG	t	TG	t	TG	t	TG	t	TG	t
God	4.30	baffled	2.76	enchanted	2.51	apparently	2.39	rot	2.27
eyes	4.21	expected	2.76	whereupon	2.51	nape	2.39	wharves	2.27
with	4.19	piety	2.75	mercy	2.50	soul	2.39	of	2.27
past	4.02	carriages	2.74	vague	2.50	wife	2.39	saints	2.27
upon	3.86	lately	2.70	insulted	2.50	imploring	2.38	myself	2.26
flesh	3.85	throat	2.69	cannot	2.49	San	2.38	floor	2.25
whom	3.75	intensely	2.68	dreadful	2.49	limply	2.37	comedy	2.25
not	3.63	cracked	2.68	toward	2.49	passion	2.36	apparel	2.22
face	3.60	perhaps	2.66	perceived	2.49	mocking	2.35	hour	2.22
mystic	3.47	remembered	2.66	awed	2.49	behold	2.35	youth	2.21
merely	3.43	lovers	2.65	open	2.48	pleasure	2.34	departing	2.20
though	3.40	amazement	2.65	planted	2.47	amidst	2.34	gravely	2.20
swiftly	3.39	while	2.65	ripped	2.47	desired	2.34	memory	2.20
whirled	3.38	senora	2.64	shops	2.47	framed	2.33	feeling	2.19
carriage	3.38	accepted	2.63	vitality	2.47	skirts	2.33	men	2.19
peace	3.36	lamps	2.62	aside	2.47	nakedness	2.33	windows	2.19
honest	3.29	moist	2.61	roofs	2.47	canes	2.32	clothing	2.19
gesture	3.22	pirates	2.61	sober	2.47	scurrying	2.32	remain	2.19
one's	3.20	damned	2.60	straightway	2.47	before	2.32	splendid	2.19
against	3.09	worrying	2.60	embarrassed	2.47	warrior	2.32	stumbling	2.19
save	3.08	decent	2.60	grace	2.47	amused	2.32	walls	2.18
relaxed	3.04	cell	2.60	primitive	2.47	lawns	2.32	horses	2.18
above	3.02	crisis	2.60	leaned	2.46	Ian	2.32	young	2.18
instead	2.98	phase	2.60	hair	2.45	Christ	2.31	savage	2.18
bidding	2.97	bound	2.60	depart	2.44	dimly	2.31	eluded	2.17

Table 3 continued

TG	t	TG	t	TG	t	TG	t	TG	t
slightly	2.97	respected	2.59	judged	2.44	blood	2.31	stirring	2.17
frail	2.96	properly	2.59	manners	2.44	calm	2.31	submit	2.17
Christian	2.94	child's	2.58	regard	2.44	anymore	2.30	while	2.17
bleeding	2.93	fierce	2.57	vigorous	2.44	life	2.30	stripped	2.17
audience	2.91	sacred	2.57	shrouded	2.44	entire	2.29	torture	2.16
equally	2.89	had	2.56	frightened	2.43	jeweled	2.29	balconies	2.16
clasped	2.89	weary	2.56	juvenile	2.43	roast	2.29	times	2.15
coldly	2.89	torn	2.55	pillows	2.43	abandon	2.29	vanity	2.15
returning	2.89	conscience	2.55	concourse	2.43	intimate	2.29	playing	2.13
mood	2.87	ruefully	2.55	smiling	2.43	patience	2.29	generations	2.13
bore	2.87	desire	2.55	faithful	2.42	virtue	2.29	cluttered	2.13
vaguely	2.87	pursuit	2.55	drops	2.42	restored	2.29	planned	2.13
assumed	2.86	thought	2.53	parting	2.42	core	2.29	started	2.13
till	2.86	moveless	2.53	religion	2.42	streets	2.29	nodded	2.12
instantly	2.84	processions	2.53	stuffed	2.42	beneath	2.28	square	2.12
dona	2.84	fires	2.53	whose	2.42	among	2.28	bowing	2.12
flowing	2.83	crossing	2.52	everybody's	2.42	infinite	2.28	candlelit	2.12
blazed	2.81	English	2.52	comers	2.41	gods	2.28	knee	2.12
gay	2.80	surging	2.52	forgive	2.41	grasped	2.27	wayside	2.12
priest	2.80	apartment	2.52	bones	2.41	bloody	2.27	dignity	2.12
agonized	2.79	recalled	2.52	leaping	2.41	self	2.27	possessed	2.12
hell	2.78	strain	2.52	witness	2.40	action	2.27	pursued	2.11
power	2.77	tumult	2.52	thrust	2.40	fleet	2.27	jerked	2.11
unfortunately	2.77	jazz	2.52	frame	2.40	increased	2.27	jaw	2.11

Table 3 continued

TG	t	JQ	t	JQ	t	JQ	t	JQ	t
rattled	2.10	you'll	-4.76	asleep	-2.87	need	-2.46	eating	-2.25
pockets	2.10	you're	-4.55	joy	-2.86	strong	-2.45	goodbye	-2.25
tongues	2.10	don't	-4.14	live	-2.83	came	-2.41	cried	-2.22
forward	2.09	can't	-4.05	turn	-2.78	morning	-2.38	happily	-2.22
mere	2.09	it's	-4.05	wouldn't	-2.76	isn't	-2.38	hoping	-2.21
quietly	2.09	that's	-3.75	tree	-2.75	leading	-2.37	Okay	-2.21
bone	2.08	get	-3.64	there	-2.75	pick	-2.36	nobody	-2.20
but	2.08	I'll	-3.53	got	-2.74	it	-2.36	we've	-2.18
aloud	2.08	let's	-3.51	ran	-2.73	ask	-2.35	every	-2.17
refused	2.08	back	-3.47	next	-2.68	neither	-2.33	pointing	-2.16
		won't	-3.42	again	-2.63	gets	-2.33	we're	-2.15
		we'll	-3.24	big	-2.62	army	-2.33	brightness	-2.15
		bring	-3.16	start	-2.61	wake	-2.32	keep	-2.15
		flew	-3.15	woke	-2.60	find	-2.31	rich	-2.15
		I'm	-3.11	just	-2.60	rushed	-2.30	baby	-2.14
		can	-3.03	doesn't	-2.59	asked	-2.30	put	-2.13
		said	-2.96	be	-2.56	couldn't	-2.30	sleep	-2.12
		after	-2.94	what's	-2.55	didn't	-2.28	weren't	-2.10
		top	-2.94	went	-2.50	liked	-2.27	passed	-2.10
		sighed	-2.93	right	-2.48	lots	-2.25	off	-2.09
		began	-2.93	poor	-2.48	how	-2.25		

"thought," "insulted," "embarrassed," "mocking," "despised," "stumbling," "stirring," "stripped," "refused," "ruefully."

Words related to time, youth and old age, departure and return in TG. Archaic and literary words like "thy," "straightway," "behold" are prevalent in the *TG* stories. Words that bring us to the nostalgic past or an awareness of time itself also predominate, such as "past," "carriage," "carriages," "memory," "times," "generations," "lately," "phase," "primitive," "youth," "child's", "young," "juvenile," "canes," "before," "hour," "eluded," "started." Words of departure and return, too, discriminate the *TG* stories: "depart," "departing," "parting," "returning," "recalled," "remembered," "remain."

Words of contrast in TG. With the use of such contrasting themes as pagan vs. Christian, darkness vs. light, weakness vs. strength, it is understandable that words of contrast also differentiate the *TG* stories from *JQ*: "not," "though," "against," "save," "instead," "cannot," "aside," "while," "but."

Examining these observations, we see that how *TG* differs from *JQ* in terms of word content fits into one harmonious picture, which Casper (1962) expresses quite well:

> the past which Joaquin actually counts is not always Christian or cosmic, but pagan and primordial. Constantly there is an urge for physical regeneration, renewal of youth, which rarely seeks sublimation or spiritual consummation. . . . Joaquin seems to be repelled by his own fascination with this brute world and its cults; or seems convinced that to embrace a religious view of this world is to accept the implications of sin as well. (p. 22)

What is most interesting here is that a word content differentiation analysis of two groups of works by supposedly the same writer reveals that the very idiosyncrasies of the writer, which earlier critics have identified again and again, have themselves become those characteristics that differentiate the writer's earlier collection from the later one. This can only mean that in the later collection the old Joaquin vanishes.

If the old Joaquin vanishes in *JQ*, what kind of Joaquin is speaking in *JQ*? An examination of the words in *JQ* that are used differently in *TG* will answer this question (see Table 3 again).

Preponderance of contracted words in JQ. JQ's biggest difference with *TG* appears to be *JQ*'s high level of informality, reflected in the following contractions: "you'll," "you're," "don't," "can't," "it's," "that's," "I'll," "let's," "won't," "we'll," "I'm," "wouldn't," "doesn't," "what's,"

"isn't," "couldn't," "didn't," "we've," "we're," "weren't." This level of informality does not go against the employment of much shorter, simpler sentences. Neither does it go against *JQ*'s low level of vocabulary richness. All these are in consonance with *JQ*'s simple writing style.

Basic action words in JQ. In addition to the auxiliary verbs "can" and "be," basic verbs that predominate in *JQ* include "get," "gets," "got," "bring," "start," "ran," "went," "came," "pick," "put," "keep," "find," "sleep," "need," "ask," "asked," "live," "turn," "said," "sighed," "began," "cried," "passed," "eating," "hoping," "pointing." These verbs are basic not only in the sense that they are understood by all grade schoolers but also in the sense they are the kind of action words that most, if not all, children can do. On the other hand, the verbs used most differently in *TG*—to cite but a few, "relaxed," "assumed," "agonized," "baffled," "forgive," "imploring," "restored," "intoned," "eluded"—presuppose a certain level of maturity.

Words of position in JQ. Words like "back," "after," "top," "there," "next," and "right" reflect the storyteller's desire to be concrete, his effort to paint a more vivid picture of the setting.

"Rich" and "poor" in JQ. While this may not constitute a preoccupation in *TG*, Joaquin in *JQ* often talks about rewarding the humble and hardworking poor with fortune and happiness at the end of his stories, the typical ending of most fairy tales.

These observations only support the fact that *JQ* was written for a young audience. Joaquin's voice in *JQ* is unprecedentedly friendly and familiar. As Binongo (1993a, 482) has noted, despite its difference from previous collections, *JQ* provides its reader with entertainment, "especially because the stories not only illustrate familiar motifs acted out by familiar characters but also employ familiar plots situated in familiar settings." Almost totally devoid of Joaquin's characteristic fixation with the Spanish past nor of his recurrent predisposition to moral and religious issues (or the absence thereof), the *JQ* collection is indeed, to use its subtitle—a Joaquinization of an originally French expression, "a la mod."

Conclusion

Our quantitative analysis supports the qualitatively obvious difference in writing style between the stories of the two collections. In terms of sentence length, the stories in *Joaquinesquerie* do not contain

any of those kilometric sentences which every *Tropical Gothic* story has. Measures of vocabulary richness also confirm the observation that the *Tropical Gothic* stories are vocabulary richer. Word content differentiation analysis reveals an even stronger result: the *Joaquinesquerie* stories reflect a lexical repertoire incompatible with traditional Joaquin's. In other words, 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.

It would be incorrect to say, however, that there are absolutely no similarities between the two collections. A patient reader could be rewarded with a few rather elusive similarities, such as those mentioned in Binongo (1993a, 478). Appearing to constitute mere coincidental occurrences rather than admitting a general tendency, these similarities, however, cannot but be submerged in the vast differences in writing style between the two collections.

A corollary to our findings is that 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. The quantitative tools that we have used cannot but point to a stylistic gulf that divides the *Tropical Gothic* and the *Joaquinesquerie* stories. For a writer hardly monolithic as Nick Joaquin, the stylometrist has to do a better job. He may be better off focusing on those very basic elements of language that a writer has to use constantly and unconsciously if he is to compose something from nothing. When a stylometrist succeeds in demonstrating that the two collections are products of the same pen, we would begin to see that quantitative analyses can do more than spectacularly demonstrating the obvious. We would be able to appreciate further the relevance of quantification to literary analysis.[10]

Notes

1. In a separate paper, Binongo (1994) shows that the difference in vocabulary richness between *TG* and *JQ* is wide enough to accommodate the vocabulary curves of Bienvenido Santos, N. V. M. Gonzalez, Edilberto K. Tiempo, and F. Sionil José.

2. Baker uses the word "pace" to refer to the type-token ratio. This is because his understanding of the term "type-token ratio" is not the ratio of types to tokens but the ratio of tokens to types. In this article, we follow the conventional practice: type-token ratio, a quantity less than or equal to unity, is the number of types divided by the number of tokens—in short, the rate at which an author generates new words.

3. A similar computation performed on the Elizabethan text data Baker provided (pp. 38-9) also suggests, although this time more strongly, that a text's type-token ratio depends on its length. The obtained r value is -0.73; with z at -8.2, the negative relationship between text length and type-token ratio is significant. The computation of Baker's r here uses only 83 of the 88 manuscripts, as data on the five texts, "Romeo and Juliet," "John Kent," "Henry VI Pt. I," "Pasquil," and "Letter to Queen," reveal computational inconsistencies, suggesting errors in input.

In an effort to correct Baker's errors, Milic (1989), without going into statistical computations, provided tables of other writers' type-token ratios to show that the type-token ratio is "inversely proportional to the size of the text." Smith (1990) added: "Baker's hypothesis can be seen false without reference to external data," and concluded his examination of Baker's proposal rather incisively: "scholars should not accept on trust a paper published in a reputable journal even when it attracts unqualified support from an apparently reliable quarter" (pp. 235-7). Holmes (1993) agreed: "[Baker's] work has no statistical theory to back up [his] contention" (p. 8).

4. The text length variance of TG is over 84 million; that of JQ is only a little over 1 million. This wide difference flatly rejects the homogeneous variance assumption and thus prevents us from using a pooled t-test.

5. Again, the t' statistic is used. The obtained t' value is 0.28, making the difference 23.48%-22.84% = 0.64% not significant at all.

6. To reduce computational burden and ease the analysis, the number of types (and the type-token rate) is monitored not after every word, but after every 25 words.

7. Honoré's R is defined by $R = (100 \log N)/(1 - V1/V)$, where $V1$ is the number of words which occur once in the text, the *hapax legomena*.

8. Because a test on the variance fails to reject the hypothesis that the variances of Honoré's R of the two collections are homogeneous, the pooled t-test is used. A t-value of 4.59 is obtained, making the result significant at 0.5% level.

9. Since there are more than 12,000 words in the combined TG and JQ corpus, we limit our study in this section to those words that occur at least 4 times. An examination of Table 3 would reveal more words that are used in TG than in JQ. Again, this is indicative of the higher level of vocabulary richness of the TG stories.

10. The present researcher has actually succeeded in showing that the two collections are products of the same pen. The results were announced in Binongo (1993b, 1993c) and are to be published in "Joaquin's *Joaquinesquerie, Joaquinesquerie's* Joaquin: A Statistical Expression of a Filipino Writer's Style" (Binongo, 1994). Essentially, a multivariable statistical technique called "principal components analysis" is employed to analyze Joaquin's usage of 36 highest frequency words: "the," "and," "to," "a," "of," etc. The application of this technique reveals that when his works are juxtaposed with those of Bienvenido Santos, N. V. M. Gonzalez, Edilberto K. Tiempo, and F. Sionil José, we see that "Joaquin . . . hardly changes the identity of the thirty-six word-types over the forty years of his mature literary career," regardless of the writing style he adopts and regardless of the stage of his growth as a writer. Thus Binongo (1994) concludes: "Indeed, an analysis based on the thirty-six most common word-types manifests the quiddity of Joaquin's writing style, clearly countering the often-overheard position that everything but change changes."

References

Ariew, R. A. 1978. André Breton's "Poisson soluble." *ALLC Bulletin* 6:34–41.

Baker, John Charles. 1988. Pace: A test of authorship based on the rate at which new words enter an author's text. *Literary and Linguistic Computing* 3:36–9.

Binongo, José Nilo G. 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, 19 May.

_____. 1993c. A statistical way of discriminating Nick Joaquin's voice in his writings. Paper read at the 20th Annual Conference of the Mathematical Society of the Philippines, 31 May.

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

Bresnahan, Roger J. 1990. *Conversations with Filipino writers*. Manila: New Day.

Burrows, J. 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*, ed. Christopher Butler. Oxford: Blackwell.

Casper, Leonard. 1962. *Modern Philippine short stories*. New Mexico: University of New Mexico.

_____. 1966. *New writing from the Philippines*. New York: Syracuse University.

Cochran, William G. and Gertrude M. Cox. 1957. *Experimental designs*. 2d ed. New York: John Wiley & Sons.

Furay, H. B. 1953. The stories of Nick Joaquin. *Philippine Studies* 1: 144–54.

Galdon, Joseph A. 1976. *Tropical gothic*: Nick Joaquin revisited. *Philippine Studies* 24:455–63.

Guerin, Wilfred L., et. al. 1992. *A handbook of critical approaches to literature*, 3d ed. Oxford: Oxford University Press.

Holmes, D. I. 1991. Vocabulary richness and the prophetic voice. *Literary and Linguistic Computing* 6:259–68.

_____. 1992. A stylometric analysis of Mormon scripture and related texts. *Journal of the Royal Statistical Society* 155:91–120.

_____. 1993. Authorship attribution. Research and consultancy report, Department of Mathematical Sciences, University of the West of England, Bristol.

Joaquin, Nick. 1952. *Nick Joaquin: Prose and poems*. Manila: Graphic House.

_____. 1972. *Tropical gothic*. St. Lucia, Queensland: University of Queensland.

_____. 1979. *Pop stories for groovy kids*. Manila: Mr. and Mrs. Publishing Co.

_____. 1983. *Joaquinesquerie: Myth a la mod*. Manila: Cacho Hermanos, Inc.

Lacaba, Emmanuel A. F. 1968. Winter after summer solstice: The later Joaquin. *Philippine Studies* 16:381–90.

Matthews, Robert A. J. and Thomas V. N. Merriam. 1993. Neural computation in stylometry I: An application to the works of Shakespeare and Fletcher. *Literary and Linguistic Computing* 8:203–9.

Milic, Louis T. 1989. A comment on John Baker's article. *Literary and Linguistic Computing* 4:153–54.

Oloroso, Laura S. 1967. Nick Joaquin and his brightly burning prose works. In *Brown Heritage: Essays on Philippine cultural tradition and literature*, ed. Antonio G. Manuud. Manila: Ateneo de Manila University.

Reuter, James B. 1956. Roots, sunlight, and rain. *Philippine Studies* 4:459–63.

Roseburg, A. G., ed. 1966. *Pathways to Philippine literature in English*. Manila: Phoenix Publishing House, Inc.

Smith, M. W. A. 1990. Attribution by statistics: a critique of four recent studies. *Revue, Informatique et Statistique dans les Sciences humaines* 26:233–51.

Tallentire, D. R. 1972. An appraisal of methods and models in computational stylistics, with particular reference to author attribution. Ph.D. thesis, University of Cambridge.

Van Peer, W. 1989. Quantitative studies of literature: A critique and an outlook. *Computers and the Humanities* 23:301–7.

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

Yule, G. Undy. 1944. *The statistical study of literary vocabulary*. Cambridge: Cambridge University.