# TWO COMPUTERIZED STATISTICAL-LINGUISTIC
# TESTS CONCERNING THE UNITY OF ISAIAH*

Y. T. RADDAY

TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY
HAIFA, ISRAEL

THE critical school of biblical studies asserts that, besides use of arguments from theology and history, the disunity of the Book of Isaiah can also be proved by the differences among its various parts in style and language. Accordingly, distinctions between First and Second Isaiah have been termed emphatic[1] and the book characterized as "an incoherent succession of ecstatic shouts,"[2] and labeled by other similar value-judgments. The subjectivity in these labels is well expressed in the comment by Lias who quoted a handbook which proclaimed that "the difference [in language between the two Isaiahs] is one to be felt rather than to be described," and then asked, "*Felt by whom?*"[3]

Many critics, moreover, have not realized the difference between style and language. The former pertains to aesthetics and is apt to be thematically conditioned and, to no small extent, left to the author's arbitrary choice, while language, since F. de Saussure[4], is, in its actual manifestation as *la parole*, governed by an author's past, reading, education, etc., and therefore completely beyond his conscious control and is by definition an idiolect and inimitable. Differences in style, even if objectively established, are consequently not author-specifying, and thus cannot serve as an argument for or against the unity of Isaiah.

[The readers for the Editorial Board of *JBL* believe that this study, worked out as part of a doctoral dissertation at Hebrew University, is significant for its methodology, even though the results provide little new in conclusions and accord with widely held critical opinion; and that, even though certain aspects of the investigation are not dealt with fully in this initial report, prompt scholarly publication is desirable so that other experts can examine the bases for findings reported in the New York Times, March 30, 1970, and elsewhere in the public press. — Editor.]

[1] Cf. A. Bentzen, *Introduction to the OT*, Vol. 2 (1949), p. 107; S. R. Driver, *An Introduction into the Literature of the OT* (1922), pp. 238 ff.

[2] R. H. Pfeiffer, *Introduction to the OT* (rev. ed., 1948), p. 465.

[3] J. J. Lias, "The Unity of Isaiah," *Bibliotheca Sacra*, 72 (1915), p. 586.

[4] F. de Saussure, *Cours de Linguistique Générale* (1916).

When the lexical inventory of the book was investigated by Cheyne[5] and Kraus,[6] their (not identical!) lists of words allegedly occurring exclusively after ch. 40 were questioned by Muilenburg,[7] himself of the school which divides Isaiah into several parts. But here too, even if these short lists (mostly consisting of only twenty to thirty words) were correct, the nonoccurrence of certain words in the first half of Isaiah could again be easily explained by the fact that the topics dealt with in the second half required the words. Moreover, some of these words occur only once after ch. 40, and there is, statistically speaking, little difference between such a word and one not occurring at all.

For similar reasons, attempts by Luzzatto,[8] Kaminka,[9] and Margaliot[10] to prove the unity of the book, by, for instance, the expression קְדוֹשׁ־יִשְׂרָאֵל recurring throughout its whole length and almost nowhere else in the Bible, and by other stylistic affinities, have to be rejected, for such expressions can be attributed to quotation,[11] imitation, or subconscious influence of one author on the other.

The only reliable author-specifying characteristics, as was already alluded to above, are those which are not governed by free choice, i.e., those of which the author is not aware himself because of their subtlety. Such characteristics are sentence length, word length, syllable entropies, transition frequencies between various parts of speech, and the like. Techniques of this kind for identification of controversial authorship have already been applied to, among other works, *De Imitatione Christi*,[12] the epistles of Paul,[13] and the epistles of Plato.[14] I hope to be able myself to publish in the near future a full report on my own similar investigations of Isaiah. It is obvious that the frequencies of language characteristics such as those mentioned cannot be calculated except by means

---

[5] T. K. Cheyne, *Introduction into the Book of the Prophet Isaiah* (1895), pp. 255–70.

[6] S. Kraus, *Sefer Yᵉšaʻyahu* (1905), p. xi.

[7] J. Muilenburg, "The Book of Isaiah," *IB*, 5 (1956), pp. 382 ff.

[8] S. D. Luzzatto, *Kerem Ḥemed*, VII, 10 (1853), pp. 224 ff.

[9] A. Kaminka, *Meḥqarim ba-Miqra uva-Talmud*, I (1935), p. 52.

[10] R. Margaliot, *Eḥad haya Yᵉšaʻyahu* (1944), passim.

[11] As regards quotation in biblical literature, see R. Gordis, "Quotations as a Literary Usage in Biblical, Oriental and Rabbinic Literature," *HUCA*, 22 (1949), pp. 157–219.

[12] G. U. Yule, "On Sentence-length as a statistical Characteristic of Style," *Biometrika*, 30 (1939), pp. 363–90. Yule's fuller examination of *De Imitatione Christi*, entitled *The Statistical Study of Literary Vocabulary* (1943), is summarized in Bruce M. Metzger, "A Reconsideration of Certain Arguments Against the Pauline Authorship of the Pastoral Epistles," *ExpTimes*, 70 (1958–59), pp. 91–94, on which, see also K. Grayston and G. Herdan, "The Authorship of the Pastorals in the Light of Statistical Linguistics," *NTStud*, 6 (1959–60), pp. 1–15.

[13] A. Q. Morton and J. McLeman, *Christianity and the Computer* (1964), and subsequent works by Morton.

[14] J. Thiele, "Einige statistische Daten der platonischen Briefe," *Grundlagenstudien aus Kybernetik und Geisteswissenschaft*, 6 (1965), pp. 109–12.

of an electronic computer. A beginning of this new methodological approach has already been suggested by Weil.[15]

In the course of my work on Isaiah, I was struck by the fact that word length, measured in syllables as defined in traditional Hebrew grammar, seemed to vary in chapter groups. In order to verify this, I drew from each chapter group three stratified samples of 500 words each and calculated the mean number of syllables per word ($\bar{\imath}$). Table 1 shows my results.

TABLE 1

*Mean Number ($\bar{\imath}$) of Syllables per Word in Isaiah*

| Section | I | II | III | IV | V | VI |
|---|---|---|---|---|---|---|
| Chapters | 1–12 | 13–23 | 24–35 | 40–48 | 49–57 | 58–66 |
| $\bar{\imath}$ | 2.06 | 2.02 | 2.05 | 2.13 | 2.08 | 2.13 |
| $\bar{\bar{\imath}}$ | 2.04 | | | 2.11 | | |

From Table 1 there emerges an unmistakable trend toward longer words in the second half of the book. The difference between the two halves as to word length will look less minute if one keeps in mind that Fucks[16] has shown that the divergence in mean word length between such a distinctly monosyllabic language as English and such a patently polysyllabic one as Turkish is but one syllable!

I hope to have the opportunity to prove in my full report that other syllable measures in Isaiah like the entropy, the mean length of chains of equal-syllabic words, etc., substantiate the prima facie case emerging from Table 1, i.e., that, in terms of an *exakte Literaturwissenschaft* as proposed by Fucks,[17] we may assume two different authors for chs. 1–35 and 40–66, respectively.

When I discussed these results with Professor C. Rabin, Hebrew University, Jerusalem, he suggested examining whether the greater

---

[15] G. E. Weil and F. Chenique, "Prolégomènes à l'Utilisation des Méthodes de Statistique Linguistique pour l'Étude Historique et Philologique de la Bible Hebraïque et de ses Paraphrases," *VT*, 14 (1964), pp. 344–66; G. E. Weil, "Méthodologique de la Codification des Textes Sémitiques servant aux Recherches de Linguistique Quantitative sur Ordinateur," *Bulletin N° 13 de l'Institut de Recherche et d'Histoire des Textes*, 14 (1964–65), pp. 115–33.

[16] W. Fucks, "Die mathematischen Gesetze der Bildung von Sprachelementen aus ihren Bestandteilen," *Nachrichtentechnische Zeitschrift*, 3, 7 (1956), p. 10.

[17] W. Fucks, *Nach Allen Regeln der Kunst: Versuch einer exakten Literaturwissenschaft* (1968).

word length in Deutero- (and Trito- ?) Isaiah may not have been caused, at least partly, by the latter's propensity to use inflected nouns.[18] Following this idea, I asked the computer to count for each chapter group the total number of words $N$, the number of nouns $N_N$, and the number of inflected nouns $N_{Nf}$, hoping thus to discover another linguistic characteristic distinguishing one author's subconscious language habits from those of another. I would like to be excused from elaborating here on the procedure how these sections of the book were determined by way of optimalization. Table 2 presents my findings for $N$, $N_N$, and $N_{Nf}$.

TABLE 2

*Percentage of Inflected Nouns in Isaiah*

| Section | $N$<br>Nr. of words | | $N_N$<br>Nr. of nouns | | $N_{Nf}$<br>Nr. of infl. nouns | | %<br>Perc. of $N_{Nf}/N_N$ | |
|---|---|---|---|---|---|---|---|---|
| I | 3363 | | 1413 | | 286 | | 18.5 | |
| II | 2351 | | 967 | | 161 | | 17.2 | |
| III | 2989 | | 1307 | | 198 | | 15.5 | |
| Σ | | 8703 | | 3687 | | 645 | | 17.1 |
| IV | 2747 | | 937 | | 238 | | 25.9 | |
| V | 2077 | | 783 | | 215 | | 27.5 | |
| VI | 2279 | | 945 | | 290 | | 30.6 | |
| Σ | | 7103 | | 2665 | | 743 | | 27.9 |
| Total | | 15806 | | 6352 | | 1388 | | 21.8 |

The picture obtained as shown in Table 2 is astonishingly neat: percentages of inflected nouns in chs. 1–35 and 40–66, respectively, lie very close to each other; the trend toward using them is decreasing in the first half of the book and increasing in the second one; and, on the whole, the frequency of inflected nouns after ch. 40 is greater by 63 percent than in the previous chapters. The characteristic investigated here may not be subtle enough to distinguish further between the language of this or that subsection of the two great halves, but justifies objectively, i.e., quantitatively, the drawing of a dividing line between the so-called First and Second Isaiah.

At this point it may be claimed that a disparity like the one just observed can be due to pure chance because such a disparity may occur also within one and the same author. This may doubtless be the case. The answer to this objection will depend on whether the disparity ob-

---

[18] From here on, I shall use the term "inflected nouns" instead of the longer, cumbersome phrase, "nouns with pronominal suffixes."

served is so significant that one may assume, with a certain measure of probability, that it is not due to chance alone but to a different authorship.

The science of statistics is, in fact, able to calculate to what measure the difference between two values found of the same variable for two sufficiently extended and unbiasedly drawn random samples is to be attributed to chance or not. The theory behind such calculations can be found by the nonstatistician in all handbooks of statistics written for the layman, e.g., in those written by Ferguson[19] and Moroney.[20] The relevant test is called the chi-square ($\chi^2$) test. It permits one to verify to what measure an observed distribution corresponds to a hypothetical random distribution.

The main idea of the chi-square test is, in short, the following. One assumes the so-called null-hypothesis, i.e., in our case, that the populations from which samples were drawn are identical. The test will tell us the degree of probability by which the said assumption may be correct. Probability values will fluctuate between 100.0 percent, namely probability close to certainty, and 0.0 percent, namely total improbability, close to impossibility.

These calculations were applied to the data noted in Table 2. The results of the chi-square test are tabulated in Table 3.

TABLE 3

*Probability Values of a Common Origin of Section Pairs*
*According to the Use of Inflected Nouns (in percents)*

| Section | II | III | IV | V | VI |
|---------|------|-------|-------|-------|-------|
| I | 5.00 | 0.10 | 0.50 | 0.05 | 0.01 |
| | II | 40.00 | 0.01 | 0.01 | 0.01 |
| | | III | 0.01 | 0.01 | 0.01 |
| | | | IV | 40.00 | 5.00 |
| | | | | V | 20.00 |

Table 3 thus proves not only — like the last column in Table 2 — that there exists much resemblance among the first three sections in Isaiah as to the use of inflected nouns on one hand and among the second three sections on the other, but how great the affinity or discrepancy is.

[19] G. A. Ferguson, *Statistical Analysis in Psychology and Education* (1959).
[20] M. J. Moroney, *Facts from Figures* ([3]1956).

Correlations among the three sections of the first half of the book fluctu-
ate between 5 and 40 percent (with section III very far removed from
section I); the same amplitude of fluctuations is to be found in correla-
tions among the three sections of the second half. All other crosswise
comparisons tell us that the chances that the sections compared originated
in the same language population are practically nil. In view of this test,
one has to assume at least one other author for the second half of the
book.

The results with these two short experiments as regards the language
behavior of various sections of the Book of Isaiah are hardly enough,
at first sight, to warrant a paper, no matter how short, since scholars
have long ago recognized the differences in these sections. Quite apart
from the fact that there still exists a small, but stubborn, minority of
those who are convinced of the unity of the book — to whom I belonged
myself before I began my own statistical work on Isaiah — there are a
few aspects of the present investigation which seem to justify the pub-
lication of this paper, although it is only a small part of a major research
on the same lines.

It is impossible to compare without measuring. The main problem
in cases of disputed authorship is not whether two texts are "alike" or
not, but in what respect and to what extent they resemble each other
or not; after all, any two texts are similar and dissimilar at the same
time. Before pronouncing a verdict of heterogeneity, not only the texts,
but the differences must be compared. Only then will one be able to
tell, by means of statistics, how high the probability is of two texts hav-
ing been written by the same person or not. This is exactly what I tried
to do here.

Further, any question of *paternitas incerta* in biblical books is of such
paramount importance not merely to researchers but in the everyday
life of most people, that one will rarely find a scholar who is sufficiently
detached and not at all *engagé* when dealing with such a topic, so as to
exclude all prejudice and subjectivity. It is, of course, irrelevant whether
Bible critics are themselves aware of this weakness, which, I admit,
they mostly are not. The approach attempted here has the invaluable
advantage of not allowing for the slightest suspicion that appraisals
were influenced by personal opinions, religious or otherwise.

Finally, from the point of view of methodology, the use of an exact
science such as mathematics for solving a problem in the humanities
answers to the *desideratum* of the ever increasing number of those who
consider the polarization of these two branches of human civilization as
one of the greatest dangers to its future.