

# McGill's Second Kierkegaard Collection

by Alastair McKinnon

*This account describes a new kind of collection consisting of the text of Kierkegaard's Samlede Værker and the software to display, search and analyse that text or any other created in similar format. It describes the creation of this text and summarizes some of the author's studies, this being a way of describing this collection and helping the reader to imagine some of the things which can be done with these tools.*

*Cet article décrit un nouveau type de collection qui comporte le texte de Samlede Værker de Kierkegaard et le logiciel qui permet de l'afficher, de la parcourir et de l'analyser ainsi que tout autre texte créé de la même façon. Il décrit la création du texte et résume certaines des études de l'auteur d'une manière qui décrit cette collection et aide le lecteur à imaginer certaines des prouesses que l'on peut réaliser avec ces instruments.*

McGill University has the rare distinction of having two quite separate and different Kierkegaard collections. The first is the Kierkegaard-Malantschuk Collection housed in the Department of Rare Books and Special Collections of the McGill University Libraries. It is a traditional collection of approximately 1,500 volumes mostly from the nineteenth century, contains a complete set of the lifetime editions of Kierkegaard's works and was acquired from the estate of the late Dr. Gregor Malantschuk, widely acknowledged as one of the world's greatest Kierkegaard scholars. This collection was the focus of a conference held at McGill in June 1980 to mark its acquisition, has been thoroughly documented in a complete catalogue published by the library in 1984 and is to be discussed in a forthcoming monograph describing some of its more important items and placing it in the context of Dr. Malantschuk's remarkable life and career. The second collection is quite new and different but Malantschuk was a man of great vision and immediately recognized its promise for Kierkegaard scholarship and, especially, for the kind of "concept analysis" in which he was particularly interested. Indeed, I have been assured, the existence of this work was the chief reason he decided that his own collection should come to McGill. There is then

no doubt that it can and should be described as "McGill's second Kierkegaard collection."<sup>1</sup>

This collection is intended primarily to foster the aims of traditional humanist scholarship but superficially is so different from traditional ones that it seems best to begin by describing it. Briefly, it consists of "PC and compatible" computer versions of the 35 works in Kierkegaard's *Samlede Værker* (Collected Works), page correlation files for each of these volumes, display and analysis software and many supplementary files to facilitate the analysis and modelling of individual concepts, books, the entire corpus, etc.<sup>2</sup> In a word, it is a computer version of these texts which can be "read" and analysed in many new and different ways. As such it would seem to be of particular interest to Kierkegaard scholars but I stress as strongly as possible that while the software is an integral part of this collection it can also be used separately, that all of it is now freely available to any member of the university for use on any similarly formatted text and that versions of all the necessary software for use with text in ASCII or free format can be provided.<sup>3</sup> Punch cards are long gone and text can now be scanned and even proof-read by machine. There is then no longer any reason why any interested person at this university should not

```

ENTEN-ELLER‡
Et Livs-Fragment
UDGIVET AF
VICTOR EREMITA‡
Første Deel‡
indeholdende A.'s Papirer‡
Er da Fornuften alene døbt,
ere Lidenskaberne Hedninger?
                                Young.‡
γ
γγγ                                Forord‡
γDet er maaskee dog stundom faldet Dig ind, kjære Læser, at tvivle
en Smule om Rigtigheden at den bekjendte filosofiske Sætning,
at det Udvortes er det Indvortes, det Indvortes det Udvortes.‡
Du har maaskee selv gjemt en Hemmelighed, om hvilken
Du følte, at den, i sin Glæde eller i sin Smerte, var Dig for kjær til
at Du kunde indvie Andre i den.‡ Dit Liv har maaskee bragt Dig i
Berøring med Mennesker, om hvilke Du ahnede, at noget Saadant
var Tilfælde, uden at dog Din Magt eller Din Besnørelse var istand
til at bringe det Forborgne til Aabenbarelse.‡ Maaskee passer intet af
Tilfældene paa Dig og Dit Liv, og dog er Du ikke ukjendt med
hiin Tvivl; den har nu og da som en flygtig Skikkelse svævet Din
Tanke forbi.‡ En saadan Tvivl kommer og gaaer, og Ingen veed,
Work EE1 Page 009 Line 13 Code M Column 01      Replace      E : 003 15

```

Figure 1. Dump of first screen of EE1.

undertake studies such as those described in this account on the author or authors of his choice. Indeed, one of the main aims of this account is to encourage him to do so.

Many readers would perhaps be puzzled at the suggestion that they might some day read books from a computer screen but a few brief comments on the "screen dump" of the opening pages of EE1<sup>4</sup> shown in figure 1 should at least give them a much better idea of the nature of this collection and some of its possibilities. The first is the copyright line which appears on the top line of the first screen but, of course, is not treated as part of the text. The next nine lines are from the title page (p. 5) of the third "base" edition.<sup>5</sup> The remaining text lines are the first 13 from p. 9, counting the title ("Forord") as the first line of this page. The cursor is not shown but it was in fact on the last line (the one beginning "*Tanke forbi*...") thus making it the "current" one. The last line in this figure is the "status" line and shows that the text displayed is EE1 and that the line

on which the cursor now rests is line 13 of page 9 of the printed text.<sup>6</sup> The "M" following "Code" indicates that this line is Main text rather than Title, Footnote, Quotation, etc. "E: 003 15" in the bottom right corner means that the counterpart of this line will be found in the then current English translation of this work on page 3, line 15.<sup>7</sup> The cent sign (¢) is used as a "sentence terminator" and inserted at all points in the text where it might be needed. One "left angle" (∧) marks the beginning of a paragraph and two or more the "level" or "nesting"<sup>8</sup> of the text beginning at that point. Thus the first two allow the computer to identify whole sentences and paragraphs and the last to divide texts at any number of specified levels, these three in order to facilitate the analysis of concepts, books, etc. About all of this, much more later.

I know of course that many humanists are or at least profess to be deeply sceptical of the value of any results produced with the aid of a computer and fear that some may even take

the above very brief account of a computer display as evidence that one ought not to take the time and effort required to understand such things. For the moment I will only say that I believe and hope that I understand these misgivings, that I began my own work with very modest aims and expectations but that I have been continually surprised as various new possibilities have opened before me. I hope that something of this sense is conveyed in the following brief account of some of the many projects which, almost fortuitously, have together resulted in the present collection.

### THE PSEUDONYMS PROJECT

Thirteen of Kierkegaard's works were written under one or more of his many pseudonyms and he attached great importance to this fact. In "A First and Last Declaration,"<sup>9</sup> appended at the last moment to what he then expected to be his last work, he pointed out that they contradicted one another, that anyone could make him look like a fool by citing one against the other, asked that anyone quoting from a pseudonymous work would "kin-

dly" do him the favour of naming the pseudonym in question and predicted that every future misunderstanding of his works would be due ultimately to a failure to take these warnings seriously. History has made him a good prophet and in 1964 I received a grant from the Canada Council to determine which of his most important pseudonyms were "closest to" or most like the acknowledged Kierkegaard. To answer this question we did a "pair-word" test based upon the theories of the late Professor Gustav Herdan using relatively large samples from eight pseudonymous and eight acknowledged works together making approximately 1,000 pages. That test showed, for example, that the ratio of observed to expected "pair words" for Anti-Climacus was 1.0087 and that for Johannes Climacus 0.8682, a result consistent with Kierkegaard's own later account of their relation<sup>10</sup> and my own earlier intuitive judgement that these two represented a "developing asymptotic approximation" to their creator.<sup>11</sup> Though flawed by an error<sup>12</sup> which no one else seems to have detected, this work was taken seriously and since that time most Kierkegaard scholars have been careful to dis-

### FRYGT OG BÆVEN

| DANSK  |               | ENG.         |  | FRAN.         |  | DEUT.         |  |
|--------|---------------|--------------|--|---------------|--|---------------|--|
| 3.udg. | 2.udg.        |              |  |               |  |               |  |
| s.     | s. l. s. l.   | p. l. p. l.  |  | p. l. p. l.   |  | s. z. s. z.   |  |
| 8      | 6 - 6         | 21 - 21      |  | 1 - 1         |  | 2 - 2         |  |
| 9      | 67 1 - 68 11  | 22 1 - 22 34 |  | 1 1 - 2 29    |  | 3 1 - 4 13    |  |
| 10     | 68 12 - 69 24 | 22 35 - 24 6 |  | 3 5 - 5 11    |  | 4 14 - 5 28   |  |
| 11     | 69 25 - 70 25 | 24 7 - 25 3  |  | 5 12 - 6 25   |  | 5 29 - 6 31   |  |
| 13     | 71 1 - 72 7   | 26 1 - 26 34 |  | 7 1 - 8 23    |  | 7 1 - 8 10    |  |
| 14     | 73 1 - 74 6   | 27 1 - 28 4  |  | 8 24 - 10 24  |  | 8 11 - 9 23   |  |
| 15     | 74 7 - 76 15  | 28 5 - 29 6  |  | 10 25 - 12 23 |  | 9 24 - 11 4   |  |
| 16     | 76 16 - 77 22 | 29 7 - 29 35 |  | 12 24 - 14 15 |  | 11 5 - 12 9   |  |
| 17     | 78 1 - 79 8   | 30 1 - 30 33 |  | 15 1 - 16 27  |  | 12 10 - 13 16 |  |
| 18     | 79 9 - 80 15  | 30 34 - 32 6 |  | 16 28 - 19 5  |  | 13 17 - 14 30 |  |

Figure 2. Excerpt from *The Kierkegaard Indices*, vol. I.

McGill's Second Kierkegaard Collection

|   | T.  | Dan.<br>3. udg. |    | Dan.<br>2. udg. |    | Eng. |    | Fr. |    | Deut. |    |
|---|-----|-----------------|----|-----------------|----|------|----|-----|----|-------|----|
|   |     | s.              | l. | s.              | l. | p.   | l. | p.  | l. | s.    | z. |
| <b>Fortvivelse</b> (Forts.)   |     |                 |    |                 |    |      |    |     |    |       |    |
| Alsaa Fortvivelse om Syndernes Forladelse er Forargelse.              | SD  | 173             | 31 | 264             | 11 | 255  | 24 | 115 | 7  | 126   | 17 |
| til det Offensive. Synd er Fortvivelse; her kjæmpes eviterende.       | SD  | 174             | 8  | 264             | 27 | 256  | 6  | 115 | 26 | 127   | 3  |
| Saa kom Fortvivelse over sin Synd; her kjæmpes endnu                  | SD  | 174             | 8  | 264             | 27 | 256  | 6  | 115 | 26 | 127   | 3  |
| Fortvivelse om Syndernes Forladelse er en bestemt Position            | SD  | 174             | 13 | 265             | 3  | 256  | 11 | 115 | 32 | 127   | 9  |
| for den Tilstand, hvori der slet ingen Fortvivelse er: i at forholde  | SD  | 180             | 10 | 272             | 2  | 262  | 26 | 122 | 4  | 134   | 5  |
| Men Angst og Frygt og Fortvivelse hjælper ikke heller til Noget.      | IC  | 72              | 28 | 86              | 18 | 69   | 40 | 176 | 30 | 65    | 34 |
| forvirrende har talt om Tvivl, hvor man skulde tale om Fortvivelse.   | IC  | 85              | 17 | 102             | 10 | 83   | 18 | 185 | 12 | 76    | 13 |
| „Fortvivelse“ viser derimod strax til Rette, ved at                   | IC  | 85              | 19 | 102             | 12 | 83   | 20 | 185 | 15 | 76    | 16 |
| talt om „Tvivl“ istedetfor at tale om „Fortvivelse“, saaledes har man | IC  | 85              | 22 | 102             | 15 | 83   | 23 | 185 | 19 | 76    | 20 |
| Alt uendelig ligegyldigt — men see, denne Sorg var Fortvivelse.       | EOT | 15              | 26 | 298             | 8  | 263  | 35 | 34  | 17 | 8     | 8  |
| hvis dog ikke til Dig; til hvem, o Fortvivelse, hvis Du virkelig      | TAF | 31              | 7  | 317             | 7  | 9    | 7  | 7   | 7  | 21    | 8  |
| straffende forfølger, da, trættet indtil Fortvivelse, at finde Hvile  | TAF | 46              | 26 | 334             | 31 | 23   | 28 | 25  | 3  | 37    | 31 |
| der dog kun fører enten dybere i Fortvivelse eller til Afsind,        | TAF | 47              | 8  | 335             | 18 | 24   | 13 | 25  | 25 | 38    | 19 |
| som maaskee indtil Fortvivelse kæmper i Haabloshed,                   | TS  | 119             | 29 | 421             | 22 | 102  | 10 | 136 | 27 | 115   | 26 |
| at ikke Forbilledet skal ængste os til Fortvivelse, bringe det i      | DS  | 184             | 31 | 499             | 3  | 171  | 7  | 205 | 12 | 187   | 4  |
| til at gaae Glip af Saligheden, eller til Fortvivelse, altsaa til at  | DS  | 213             | 25 | 533             | 14 | 202  | 1  | 236 | 19 | 221   | 18 |
| jeg maatte enten styrte mig i Fortvivelse og Sandselighed,            | SFV | 89              | 39 | 561             | 18 | 18   | 14 | 65  | 27 | 31    | 10 |
| som vilde bringe de Fleste til Fortvivelse; thi det er                | SFV | 109             | 23 | 584             | 2  | 47   | 17 | 84  | 10 | 54    | 1  |
| efter i Fortvivelse at have brudt med Verden og hvad Verdens          | SFV | 112             | 18 | 587             | 13 | 52   | 2  | 86  | 34 | 57    | 19 |
| han endnu engang sit Onske „o, Fortvivelse, det opfyldes ikke,        | O   | 275             | 13 | 316             | 13 | 244  | 13 | 141 | 35 | 285   | 27 |
| Fortvivelse, det opfyldes ikke, Fortvivelse, man lever kun            | O   | 275             | 13 | 316             | 13 | 244  | 13 | 141 | 35 | 285   | 27 |
| der, indtil Fortvivelse smerteligt, skiller En fra                    | O   | 322             | 21 | 370             | 9  | 286  | 1  | 178 | 19 | 333   | 4  |

Figure 3. Excerpt from *The Kierkegaard Indices*, vol. II.

|                            |  |                          |
|----------------------------|--|--------------------------|
| <b>ukjendelig</b> [21]     | <b>Ukjerligt</b> [1]                   | <b>Ukrudtet</b> [1]      |
| BI 1 169 9                 | KG 12 283 24                           | TTL 6 284 25             |
| T 4 266 *                  | <b>ukjerligt</b> [15]                  | <b>Ukrudts</b> [1]       |
| G 5 181 18                 | KG 12 17 32 173 4 210 11 212 15 212 23 | LA 14 57 26              |
| TTL 6 309 11 312 13 312 24 | 244 29 245 15 267 6 280 17 283 23      | <b>ukrydret</b> [1]      |
| SV1 7 28 26                | 284 29 301 9 324 27                    | EE2 3 129 26             |
| AE2 10 277 4               | IC 16 180 36                           | <b>ukrænkede</b> [1]     |
| LT 11 17 33                | O 19 123 21                            | BI 1 191 21              |
| HLF 11 157 26 194 6        | <b>ukjær</b> [1]                       | <b>ukrænket</b> [4]      |
| LE 11 227 1                | EE2 3 145 35                           | BI 1 69 14               |
| KG 12 16 39 41 29          | <b>ukjærlig</b> [2]                    | EE2 3 33 21 102 1        |
| CT 13 60 16                | BI 1 307 28                            | SV1 7 18 16              |
| IC 16 128 8 234 4 234 7    | EE2 3 15 26                            | <b>ukunstneriske</b> [1] |
| DS 17 227 27               | <b>uklar</b> [8]                       | EE1 2 269 15             |
| <b>Ukjendelige</b> [11]    | EE1 2 343 3                            | <b>ukyndig</b> [4]       |
| LA 14 98 6 98 28 99 *      | EE2 3 47 6 303 13 303 37               | BI 1 104 31 137 25       |
| IC 16 128 *                | BA 6 227 37                            | G 5 130 34               |
| <b>ukjendelige</b> [7]     | TTL 6 268 18 293 22                    | SV1 7 33 29              |
| AE1 9 20 2                 | KG 12 344 11                           | <b>Ukyndige</b> [4]      |
| KG 12 71 22 296 26         | <b>Uklare</b> [2]                      | BI 1 158 *               |

Figure 4. Excerpt from *The Kierkegaard Indices*, vol. III.

McGill's Second Kierkegaard Collection

| Total  | LP   | BI | EE1 | EE2 | T   | FB  | G  | F  | PS  | BA | TTL | SV  | AE  | LA  | OTA | KG  | CT  | KK | SFV | LF | TSA | SD | YTS | IC | EOT | GU | TAF | FV | TS | DS | B21 | Ø  | HCD | BFF |   |
|--|------|----|-----|-----|-----|-----|----|----|-----|----|-----|-----|-----|-----|-----|-----|-----|----|-----|----|-----|----|-----|----|-----|----|-----|----|----|----|-----|----|-----|-----|---|
| Religjøsitet (3)                             | 225  |    |     |     |     |     |    |    |     | 1  | 25  | 157 | 17  |     |     |     |     |    | 16  |    | 2   |    |     |    |     |    |     |    |    | 1  | 1   |    | 5   |     |   |
| Ret  | 564  | 12 | 46  | 25  | 81  | 27  | 16 | 19 | 5   | 2  | 7   | 4   | 37  | 57  | 6   | 50  | 38  | 30 | 4   | 7  | 1   | 7  | 18  | 2  | 20  |    | 1   |    | 6  | 16 | 7   | 11 |     | 2   |   |
| Revolutions-Tiden (2)                        | 26   |    |     |     |     |     |    |    |     |    |     |     |     |     | 26  |     |     |    |     |    |     |    |     |    |     |    |     |    |    |    |     |    |     |     |   |
| Sandhed                                      | 1508 | 12 | 52  | 47  | 108 | 118 | 4  | 9  | 5   | 12 | 20  | 14  | 43  | 220 | 10  | 202 | 109 | 48 | 6   | 37 | 16  | 25 | 14  | 27 | 147 | 1  | 5   | 4  | 4  | 18 | 54  | 18 | 91  | 2   | 6 |
| Schelling (5)                                | 18   |    | 1   |     | 1   |     |    |    |     |    | 12  |     |     | 4   |     |     |     |    |     |    |     |    |     |    |     |    |     |    |    |    |     |    |     |     |   |
| Schlegel (6)                                 | 44   |    | 42  |     |     |     |    |    |     |    | 1   |     | 1   |     |     |     |     |    |     |    |     |    |     |    |     |    |     |    |    |    |     |    |     |     |   |
| Schleiermacher (4)                           | 40   |    | 36  |     |     |     |    |    |     | 2  |     | 1   |     |     |     |     |     |    |     |    |     |    |     |    |     |    |     |    |    |    |     |    |     | 1   |   |
| Skepals                                      | 56   | 1  | 9   |     | 13  |     |    |    |     | 6  | 3   | 1   | 18  | 5   |     |     |     |    |     |    |     |    |     |    |     |    |     |    |    |    |     |    |     |     |   |
| Socrates (16)                                | 1399 | 1  | 967 |     |     | 11  | 1  | 1  | 123 | 24 |     | 43  | 157 | 6   |     | 2   |     | 12 |     |    | 5   | 45 |     | 1  |     |    |     |    |    |    |     | 7  | 3   |     |   |
| Sophisten (18)                               | 213  |    | 139 | 5   | 3   |     | 2  | 1  | 4   | 12 |     | 16  | 10  | 1   |     |     |     | 2  |     |    |     | 1  |     |    |     |    |     |    |    |    |     | 8  |     |     |   |
| Speculation (2)                              | 174  | 1  | 15  |     | 2   |     |    |    |     | 1  | 7   | 2   | 140 |     |     |     |     |    |     |    | 1   | 2  |     | 3  |     |    |     |    |    |    |     |    |     |     |   |
| Spinoza (4)                                  | 13   |    | 2   | 1   |     |     |    |    |     | 7  |     |     | 3   |     |     |     |     |    |     |    |     |    |     |    |     |    |     |    |    |    |     |    |     |     |   |
| Sprog (language)                             | 155  | 1  | 6   | 15  | 4   | 11  | 3  | 7  | 3   | 3  | 1   | 3   | 28  | 13  | 1   | 14  | 9   | 11 |     |    |     | 6  |     | 7  |     |    | 1   | 1  | 2  |    |     | 3  |     |     |   |
| Sprogbrug (usage)                            | 43   |    | 2   |     | 3   | 2   |    |    | 1   | 4  | 6   |     | 1   | 2   |     |     | 15  | 2  |     |    |     |    |     |    |     |    |     |    |    |    |     | 4  | 1   |     |   |
| Stoicisme (4)                                | 10   | 1  | 2   |     | 1   |     |    |    |     | 2  |     |     | 1   |     |     |     |     |    |     |    |     | 2  |     |    |     |    |     |    |    |    |     | 1  |     |     |   |
| Subjectivitet (6)                            | 243  |    | 62  | 5   |     | 2   |    |    |     | 10 |     | 2   | 154 | 1   |     |     |     |    |     |    |     |    |     | 5  |     |    |     |    |    |    |     |    |     |     |   |
| System (2)                                   | 246  | 3  | 20  | 2   | 4   | 10  | 1  | 19 | 4   | 8  |     | 17  | 125 | 2   |     |     |     | 13 |     |    |     | 5  | 1   |    |     |    |     |    |    | 1  |     |    | 11  |     |   |
| Tanke-Existentia (conceptual existence)      | 7    |    |     |     |     |     |    |    |     |    |     | 1   | 6   |     |     |     |     |    |     |    |     |    |     |    |     |    |     |    |    |    |     |    |     |     |   |
| Tankebestemmelse (4) (thought-determination) | 14   |    |     |     |     |     |    |    |     | 1  |     |     | 5   |     |     | 1   |     | 1  | 1   |    | 1   |    | 4   |    |     |    |     |    |    |    |     |    |     |     |   |
| Tankebevægelse (3) (line of thought)         | 7    |    | 3   | 2   |     |     |    |    |     |    |     | 2   |     |     |     |     |     |    |     |    |     |    |     |    |     |    |     |    |    |    |     |    |     |     |   |
| Tautologi (7)                                | 22   |    | 3   | 4   | 1   | 1   |    | 1  | 4   | 1  |     | 1   | 6   |     |     |     |     |    |     |    |     |    |     |    |     |    |     |    |    |    |     |    |     |     |   |
| Teleologi (6)                                | 73   |    | 3   | 1   | 26  | 9   |    |    | 1   | 1  |     | 3   | 17  |     |     |     |     | 2  |     |    | 9   | 1  |     |    |     |    |     |    |    |    |     |    |     |     |   |
| Thales                                       | 3    |    |     |     |     |     |    |    |     |    |     | 2   |     | 1   |     |     |     |    |     |    |     |    |     |    |     |    |     |    |    |    |     |    |     |     |   |
| Theaetetus (3)                               | 4    |    | 2   |     |     |     |    |    | 2   |    |     |     |     |     |     |     |     |    |     |    |     |    |     |    |     |    |     |    |    |    |     |    |     |     |   |
| Theisme (2)                                  | 2    | 1  | 1   |     |     |     |    |    |     |    |     |     |     |     |     |     |     |    |     |    |     |    |     |    |     |    |     |    |    |    |     |    |     |     |   |
| Thrasymachus                                 | 14   |    | 14  |     |     |     |    |    |     |    |     |     |     |     |     |     |     |    |     |    |     |    |     |    |     |    |     |    |    |    |     |    |     |     |   |

Figure 5. Sample page from "Kierkegaard's Remarks on Philosophy."

tinguish at least between the pseudonymous and acknowledged works.<sup>13</sup> The situation was not without its own peculiar irony. As I remarked at the end of one of my reports, "Of course, it is ludicrous that anyone should have to use a computer to show that Kierkegaard understood the structure of his authorship but, as he often remarked, where illusions are deep-seated and wide-spread, a certain kind of deceit is sometimes necessary."<sup>14</sup>

### THE KIERKEGAARD INDICES

Before I had completed the pseudonyms project I realized that the text of Kierkegaard's *Samlede Værker* already created should be used as the basis for a concordance to these works and applied to the Canada Council for funds to create the remaining 5,000 pages of text and to pay for the necessary programming. In fact this projected work finally appeared as volume II of *The Kierkegaard Indices*, the history of which is described below.<sup>15</sup>

Though I have always used the current Danish third as our "base" edition I realized that some Danish scholars preferred the second. I also recognized that Danish is not widely known outside that country, that this concordance should therefore also provide references to the current English, French and German translations and that this would require the creation of extensive page correlation files. However, once these had been created, it was immediately obvious that they would also be valuable to others and, wanting some computer typesetting experience before undertaking the concordance, we published these tables under the title *Kierkegaard In Translation: En Traduction: In Übersetzung* as volume I of the series. Figure 2 shows the first ten lines of these tables for FB. The last of these lines shows that the counterpart of all of p. 18 in the Danish third edition runs from p. 79, l. 9 to p. 80, l. 15 in the Danish second edition, from p. 30, l. 34 to p. 32, l. 6 in the English translation, etc. Thus, for example, the user of the English translation can readily

McGill's Second Kierkegaard Collection

| Title code | Corrected mean date of composition | Classification | Gud  | Christus | Jesus | Jesus Christus | Herre Jesus Christus | Gud-Mennesket | Hellige-Aand | Guden | Guder | All terms |
|------------|------------------------------------|----------------|------|----------|-------|----------------|----------------------|---------------|--------------|-------|-------|-----------|
| LP         | 1 July 1838                        |                | 0.02 | 0.01     |       |                | 0.01                 |               |              |       |       | 0.04      |
| BI         | Dec 1840                           |                | 0.05 | 0.02     |       |                |                      |               |              |       | 0.04  | 0.11      |
| EE1        | 1 Apr 1842                         | p              | 0.03 |          |       |                |                      |               |              |       | 0.02  | 0.05      |
| EE2        | 1 Apr 1842                         | p              | 0.17 | 0.01     |       |                |                      |               |              |       |       | 0.18      |
| G          | June 1843                          | p              | 0.13 |          |       |                |                      |               |              |       | 0.01  | 0.14      |
| FB         | 1 July 1843                        | p              | 0.27 |          |       |                |                      |               |              | 0.01  |       | 0.28      |
| T          | Oct 1843                           | a              | 0.59 | 0.01     | 0.01  |                |                      |               |              |       |       | 0.61      |
| BA         | Apr 1844                           | p              | 0.08 | 0.02     |       |                |                      |               |              |       |       | 0.10      |
| PS         | May 1844                           | p              | 0.05 |          |       |                |                      |               |              | 0.53  |       | 0.58      |
| F          | 1 June 1844                        | p              | 0.04 |          |       |                |                      |               |              |       |       | 0.04      |
| SV         | Dec 1844                           | p              | 0.14 |          |       |                |                      |               |              | 0.01  | 0.04  | 0.19      |
| TTL        | 1 Jan 1845                         | a              | 0.42 |          |       |                |                      |               |              |       |       | 0.42      |
| BFF        | 1842-1851                          |                | 0.08 |          |       |                |                      |               |              | 0.01  | 0.02  | 0.11      |
| AE         | July 1845                          | p              | 0.23 | 0.02     |       |                |                      |               |              | 0.03  |       | 0.28      |
| LA         | Jan 1846                           | a              | 0.03 |          |       |                |                      |               |              | 0.02  |       | 0.05      |
| OTA        | Aug 1846                           | a              | 0.49 | 0.08     |       |                |                      |               |              |       |       | 0.57      |
| KK         | Feb 1847                           | p              | 0.03 |          |       |                |                      |               |              |       |       | 0.03      |
| KG         | May 1847                           | a              | 0.42 | 0.06     |       |                |                      |               |              |       |       | 0.48      |
| TSA        | Oct 1847                           | lp             | 0.33 | 0.24     |       | 0.01           | 0.01                 | 0.01          |              |       |       | 0.60      |
| CT         | Nov 1847                           | a              | 0.84 | 0.07     |       |                | 0.01                 |               |              |       |       | 0.92      |
| SD         | Feb 1848                           | lp             | 0.56 | 0.12     |       |                |                      | 0.02          |              |       |       | 0.72      |
| IC         | Aug 1848                           | lp             | 0.34 | 0.44     | 0.01  | 0.04           | 0.02                 | 0.10          | 0.01         |       | 0.01  | 0.95      |
| SFV        | Sept 1848                          |                | 0.29 | 0.02     |       |                |                      |               |              |       |       | 0.31      |
| LF         | Mar 1849                           | a              | 1.22 |          |       |                |                      |               |              |       |       | 1.22      |
| FV         | Mar 1849                           |                | 0.24 |          |       |                |                      |               |              |       |       | 0.24      |
| YTS        | 1 Sept 1849                        | a              | 0.76 | 0.14     |       | 0.02           | 0.08                 |               |              |       |       | 1.00      |
| TAF        | Dec 1849                           | a              | 0.34 | 0.13     |       | 0.06           | 0.04                 |               |              |       |       | 0.57      |
| EOT        | Nov 1850                           | a              | 0.30 | 0.50     |       |                |                      |               |              |       |       | 0.80      |
| GU         | May 1851                           | a              | 1.04 |          |       |                |                      |               |              |       |       | 1.04      |
| TS         | July 1851                          | a              | 0.54 | 0.11     | 0.01  |                |                      |               | 0.06         | 0.01  |       | 0.73      |
| DS         | Nov 1851                           | a              | 0.43 | 0.06     |       | 0.01           | 0.01                 |               |              |       |       | 0.51      |
| B21        | Mar 1855                           | l              | 0.25 | 0.09     |       | 0.03           | 0.01                 |               |              |       |       | 0.38      |
| Ø          | May 1855                           | l              | 0.40 | 0.12     |       | 0.01           | 0.01                 | 0.01          |              |       |       | 0.55      |
| HCD        | 5 June 1855                        | l              | 0.20 | 0.66     |       | 0.07           |                      |               |              |       |       | 0.93      |

Figure 6. Relative frequencies of key theological terms.

calculate that the word translated as "impotence" on p. 31, l. 23 will appear on p. 18, l. 21 of the Danish third edition and, turning to this point in the text, find that it is actually *Afmagt*. The original computer version of these tables is the basis of all such "secondary" references in the concordance and, with the later addition of comparable information for the Danish first edition, of all such references now available on request in the bottom right corner of the computer screen.

It was also clear to me from the beginning that the key-word-in-context or so-called KWIC concordances then in fashion were quite unacceptable and that our contexts or fragments should be as coherent and perspicuous as possible. In order to achieve this goal we assigned various weights to different punc-

tion marks in the text giving the highest to the period, question mark and explanation mark and the lowest to the comma. Thus the program begins with the concordance head or key word in question, moves alternately one word in each direction until stopped by a sufficient "weight" of punctuation on either side and completes the fragment by including whole words on the other side falling within the allotted 70 characters and blank spaces. Some typical results of this approach are shown in Figure 3 which reproduces 22 of the 70 lines or fragments on each page of this concordance. The first text line shows that this particular instance of *Fortviølelse* (despair) is from SD and occurs on p. 173, l. 31 of the Danish third edition, p. 264, l. 11 of the Danish second edition, p. 255, l. 24 of the then current English translation, etc. Note that an inspec-

tion of the original texts shows that punctuation has determined one or other terminus of 16 of these 22 fragments and both termini of five. Note, too, that of the English references cited, 10 or 45.5% are completely correct, 11 or 50% accurate within one line and 1 or 4.5% accurate within two. I have not checked the other "secondary" references shown in this excerpt but note that these figures are very similar to those reported to me by others. This work, the original impetus to the series, was entitled *Fundamental Polyglot Konkordans til Kierkegaards Samlede Værker* and published as vol. II.

Ideally I would have liked to treat all important words in this rather elaborate and expensive way but the original text contains 55,568 word-types and 1,944,767 word-tokens<sup>16</sup> and this was obviously out of the question. Indeed, even the 1137 very large pages of this volume concord only 586 types together with approximately 210 variants. Accordingly we decided to provide a simple index of most of the remaining words in a work entitled *Index Verborum til Kierkegaards Samlede Værker* and published as volume III. In fact, even its 1322 equally large pages allowed us to index only the remaining words having a frequency less than 500 plus certain selected higher frequency words which play a key role in Kierkegaard's thought and other forms of which had already been treated in the concordance. The first 20 lines of p. 1199 of this volume are reproduced in Figure 4. The first shows that the word *ukjendelig* (unrecognizable) occurs 21 times in the entire text, that the first instance is found in BI (or volume 1 of the third edition according to Danish and German use) on p. 169, l. 9, the next that there are three or more occurrences of this word on p. 266 of T (or volume 4 of this edition), etc. Those using other editions or translations can of course convert these references with the help of the tables in volume I already described.

Just as I was beginning to think and hope that I was at last finished with the burden of these quite traditional and obvious works I chanced upon Kučera and Francis' *Computational Analysis of Present Day American English*<sup>17</sup> and saw immediately that

a comparable analysis of Kierkegaard's writings might prove very helpful in understanding his thought. With their encouragement I therefore proceeded to produce a counterpart volume entitled *Computational Analysis of Kierkegaard's Samlede Værker* which appeared in 1975 as volume IV in this series. I quote from the Introduction:

Even without benefit of interpretation the reader will see the relevance of the following examples all of which are taken directly from this work or can be readily constructed therefrom. Of the 541 occurrences of *absurd*, *Absurde*, *Absurditet*, *Paradox*, *paradox*, *Paradoxet*, only two are from his acknowledged works. The correlation coefficient of *Abraham* and *Isaak* for all of the works is + 0.999, which result is significant at the 0.01% level. That between the passage of time and Kierkegaard's use of a selected set of "neutral" Christian terms is + 0.889, itself significant at the 0.01% level. The ratio between Kierkegaard's use of the various forms of *Styrelse* and *Skjebne* shows significant changes at mid December, 1845 and again during the period July-November, 1848. Certain allegedly key terms in Kierkegaard's thought (e.g. *existentielle*, *Existents-Inderlighed*, *Existents-Meddelelse*, *Existentsens*, *Existerende*, *existerende*) are effectively restricted to AE, a pseudonymous work written between April and December of 1845.<sup>18</sup>

I might state some of these points differently today but remain convinced that this is the most important of these four volumes and that this will become increasingly evident as the computer text and hence custom made concordances become more widely available. Certainly it persuaded me that statistics can help us identify the distinctive features of a text and that interpretations failing to explain these features cannot possibly be accepted as adequate or true. This realization has guided

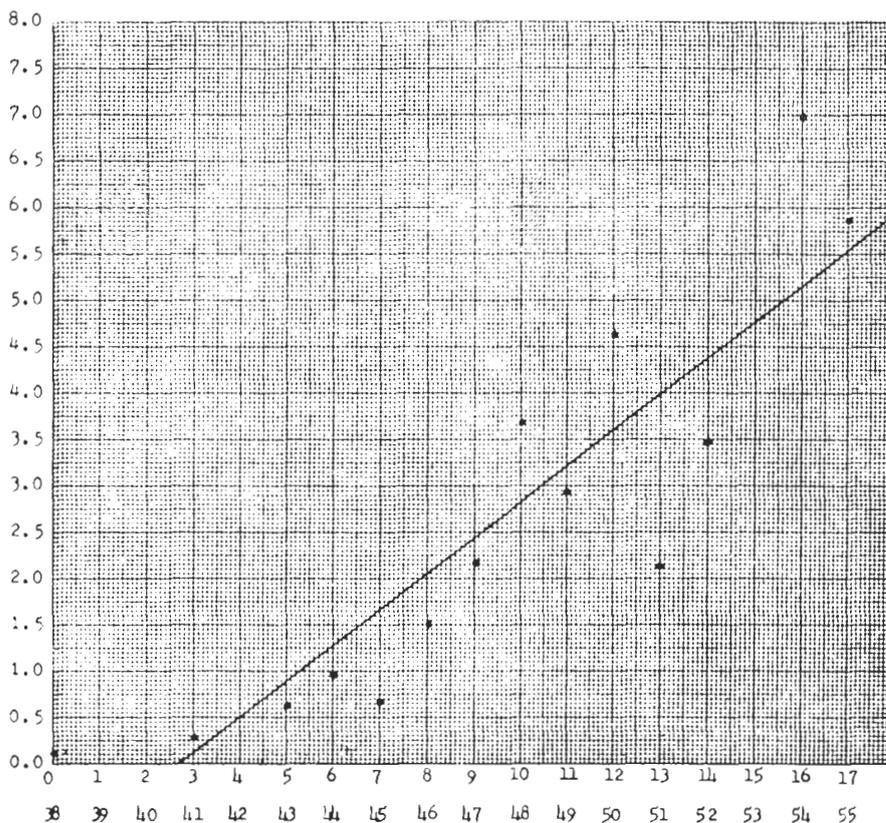


Figure 7. The increase of Christian terms across authorship.

much of my work since that time and, if more widely accepted, could greatly improve the quality of literary studies.

### TRANSITION PIECES

Of course one does not change the focus of one's work and thought overnight and the record shows that I actually did some studies inspired by Kučera and Francis as early as 1973. These were not properly statistical and were done on the mainframe using awkward and now outdated programs but they are interesting in themselves and pointed the way to others which are obviously more important and which can now be done much more easily with the software available in this collection. We therefore begin with a brief summary of a few of these early and very simple studies.

One of the reasons Kierkegaard has not been taken more seriously as a philosopher is that his many philosophical remarks and insights occur within the context of a vast and wide ranging literature which is not itself primarily philosophical, at least in the narrow sense of that term. I sought to overcome this difficulty in a short and very densely packed piece entitled "Kierkegaard's Remarks on Philosophy"<sup>19</sup> one page of which is shown as Figure 5. This data array is self-explanatory but the following brief comments should remove any possible confusion. The title codes across the top of the table represent Kierkegaard's works ordered, with one exception, according to their date of publication. The words in the left hand column represent names or topics of particular philosophical interest and the figure in brackets indicates

the number of forms included in the count. The second column represents the number of times the word or words occur in the corpus as a whole and the remaining columns the number of times they occur in the book in question. Of course the interested scholar can learn a great deal even from such an array. For example, the first row indicates that the three main forms of *Religieusiteten* (religiosity) occur 225 times in the corpus, for the first time in BA and most frequently in AE, that two main forms of *Theisme* together occur only twice and at the beginning of the authorship, etc.

A similar report dealing with another important area of his thought, "Theological Focus in Kierkegaard's *Samlede Værker*: Some Basic Data,"<sup>20</sup> was published in 1974 and shows the relative frequencies of nine key theological terms in all the works of the authorship. The core of this report is shown in Figure 6. The first column shows the title code of each work, the second its "corrected mean date of composition,"<sup>21</sup> the third its classification (pseudonymous, acknowledged, late pseudonymous), the next nine the relative frequencies of the name in the work in question and the last the relative frequencies<sup>22</sup> of all these names in that work. Note that this table cites relative frequencies in order to permit easy and meaningful comparisons between both books and names. Thus, for example, one can see immediately that *Gud* (God) is approximately three times more frequent in T than in EE2 and approximately 59 times more frequent than *Christus* (Christ) in T. Later we shall present some more recent and perspicuous graphic representations of such basic information but meanwhile note that this very simple summary provides a great deal of information in a very useful form and is an obvious model for anyone seeking to improve his own overall grasp of a particular area or to lay the groundwork for other more serious studies in the field.

Partly as a follow up of this study I published "The Increase of Christian Terms in Kierkegaard's *Samlede Værker*"<sup>23</sup> in 1974. This piece documents the increase in the average number of occurrences per page of a set of 13 "neutral"<sup>24</sup> Christian terms across the years of

the authorship. These data are plotted and their trend line shown in Figure 7. Note that there are significant "drop-offs" from previous levels in the years 1851 and 1855 and that the correlation coefficient between the frequencies of this sample set and the years in question is + 0.889 and hence so high that it clearly requires explanation. This study involved a great deal of manual counting and calculation but most of the programs for similar studies giving even more accurate results are now available in our collection.

"Kierkegaard's Literary Production by Quarterly Rates"<sup>25</sup> shows the approximate number of words from the *Samlede Værker* and the present three different "groups" of the *Papirer* (Kierkegaard's unpublished writings) written during each quarter of the years 1834 to 1855 in both tabular and graphic form. The graph for the *Samlede Værker* is shown in Figure 8 and that for all writings combined in Figure 9. Note that most of the early writings are from the *Papirer* and that his authorship proper begins around mid-1840, peaks in 1842, 1844-45 and again in the third quarter of 1847, virtually ceases in 1850 and thereafter accounts for more than half of his total production only during the second quarter of 1855. This information is particularly valuable in view of the fact that he seems to have been almost driven to write and, indeed, to publish. It tells us a great deal about the connection between his life and work and provides a background against which these and many other questions may be answered more precisely. Presumably similar studies would also be useful for most figures with whom humanists are concerned.

"A Method of Displaying Differences between Various Accounts of an Object"<sup>26</sup> is from a time when I had recognized the importance of identifying the distinguishing features of at least three different accounts of the same object but had not yet refined any of my present techniques or devised any of my present software for doing so. In fact it employs some relatively crude criteria, the KYST<sup>27</sup> multi-dimensional scaling program and a simple formula to convert its results to a "pseudo three-dimensional" space, but nevertheless clearly succeeds in identifying those words which dis-

McGill's Second Kierkegaard Collection

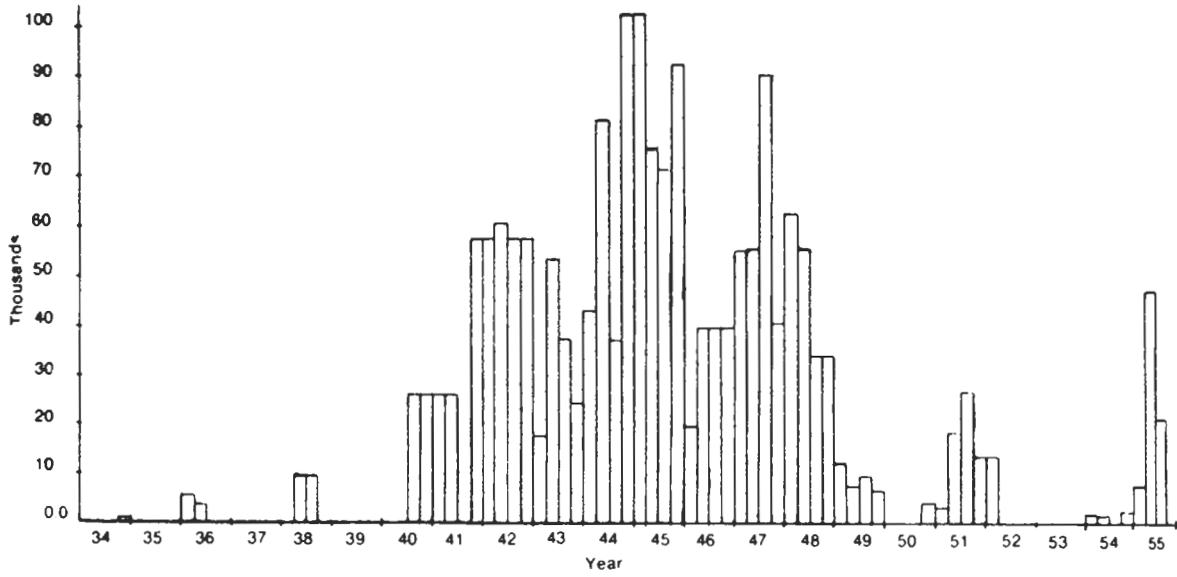


Figure 8. Kierkegaard's quarterly production: *Samlede Værker*.

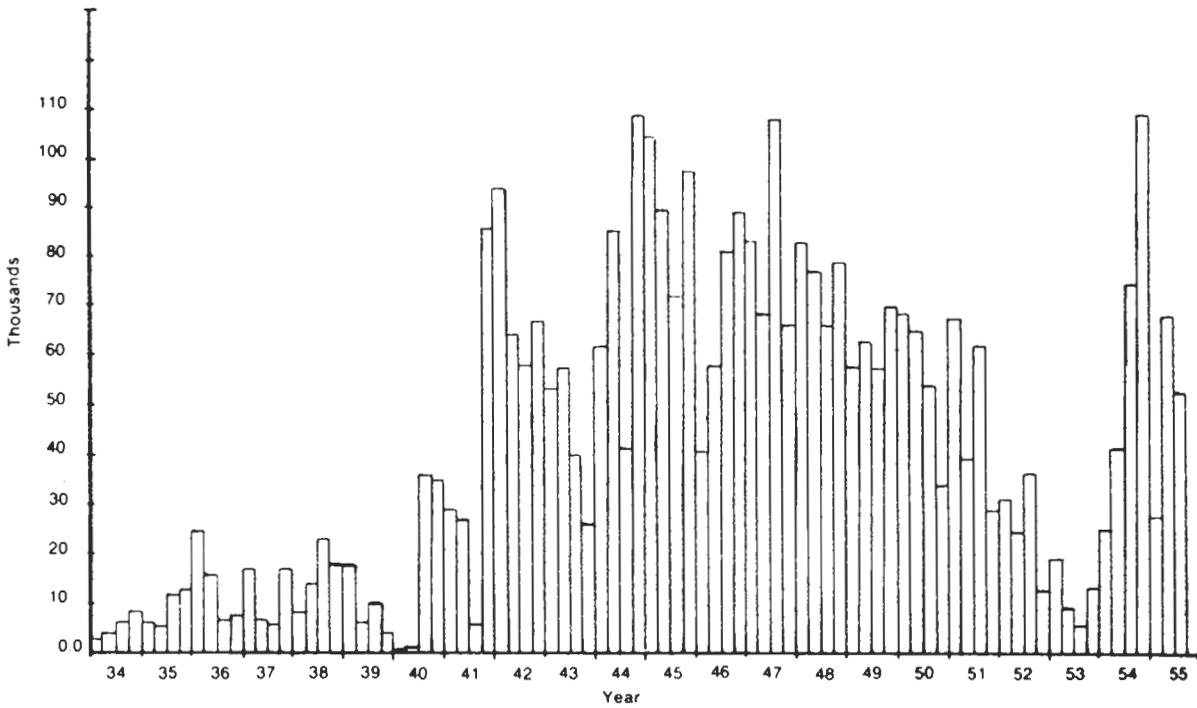


Figure 9. Kierkegaard's quarterly production: all writings.

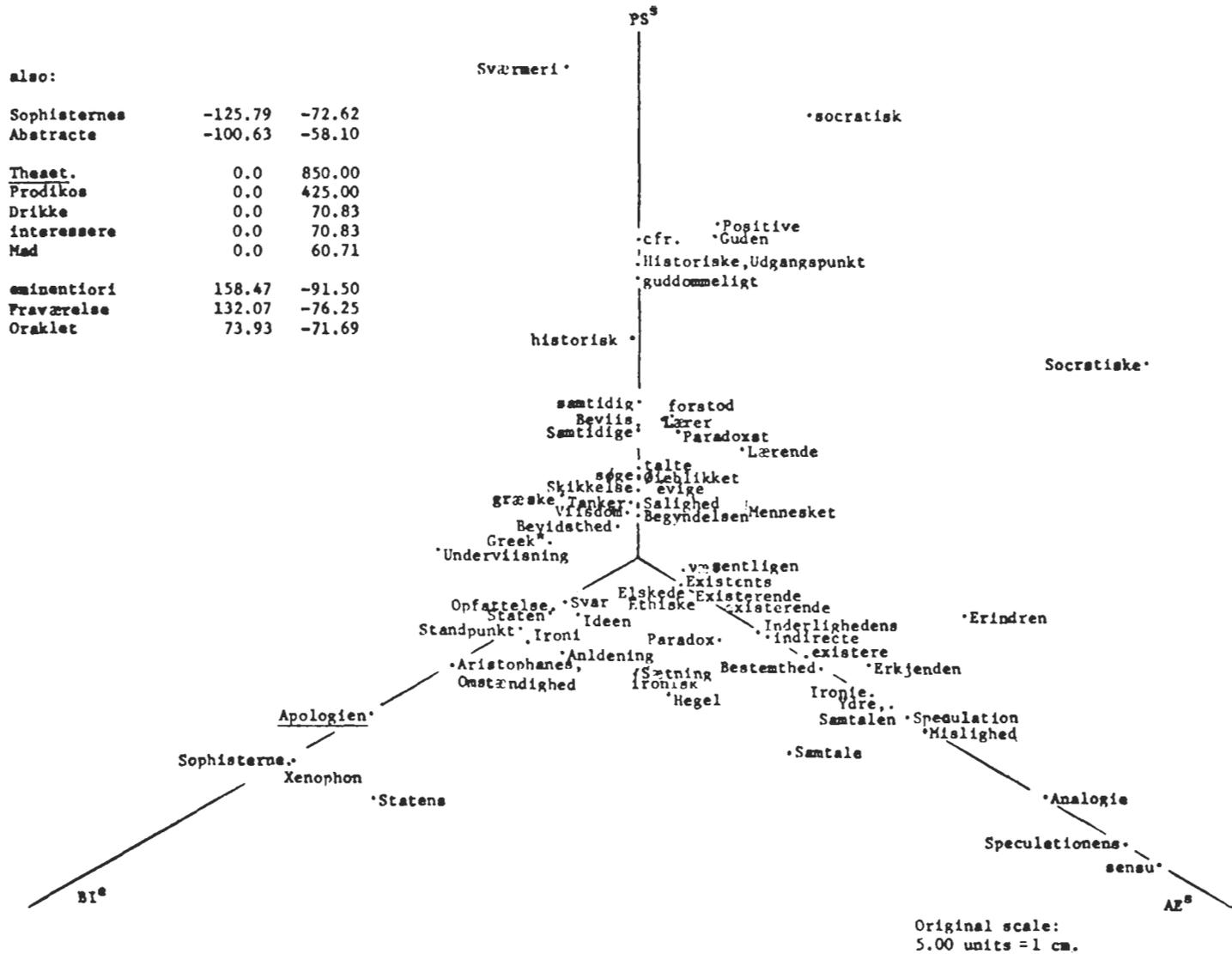
tinguish Kierkegaard's accounts of Socrates in BI, PS and AE, the three works in which he figures most prominently. The summary results are given in Figure 10 and, translated, show that Socrates is strongly and uniquely associated with Xenophon, the Sophists, *The Apology*, Aristophanes, *The Republic*, etc. in BI, with *Theaetetus*, Prodicus, the historical, divine, contemporary, proof, contemporaneity in PS and with *sensu*, speculation, analogy, misunderstanding, conversation, appearance, irony (with a final "e" in the original), certainty, to exist, etc. in AE. Note that they also show that, for example, he is clearly associated with Hegel in both BI and AE and with Socrates, the Socratic and recollection in both PS and AE. I believe that anyone who knows these works well will agree that these words point to the distinguishing characteristics of their different accounts of Socrates and that they will therefore enable one to grasp and understand these differences. I also believe that this approach could help us understand other important authors presenting similarly evolving accounts of some subject or figure. Note however that, as we shall see, our collection includes newer programs capable of producing much more detailed, sophisticated and perspicuous results which can be used for this same purpose.

Kierkegaard made constant reference to and was greatly influenced by the Bible and it is therefore important to document his perception of it as precisely and accurately as possible. I attempted to do this in "Kierkegaard's Perception of the Bible"<sup>28</sup> the results of which were originally generated by KYST but are shown here in part in a more recent three-dimensional plot as Figure 11. Note, for example, that Mark, Luke and Matthew are the most central, that most of the books of the Old Testament lie on the right of this space, etc. These results are entirely consonant with and, indeed, help us understand his writings. Had they been produced by our present software, they would contain at least as much information as, say, 10 pages of text in, and this is very important, a readily memorable form. I have no doubt that similar results for many other authors and their sources would be equally instructive and valuable.

## SOME NECESSARY BACKGROUND

In 1981 I moved my Kierkegaard text from the mainframe to my own PC and began thinking about the creation of software which would produce better and more perspicuous results, emphasise exploratory data analysis rather than classical statistics, make more use of graphic representations and, above all, explore the possibilities of representing conceptual relations in multi-dimensional space, the last because I saw even then that this would be the next real frontier for computers and, for that matter, our understanding of text.<sup>29</sup> Of course this is and should be a continuing process but more than thirty of these programs have been completed and are included in the collection under discussion. Most of these presuppose a number of basic notions such as aberrant frequency word, data matrix, word profile, change point detection, correspondence analysis and the admittedly puzzling but, I believe, helpful and very promising idea of the "space" of a text. I shall now attempt to explain and illustrate these notions in terms of a book and its parts but stress that this is simply a matter of convenience and that these notions apply equally to any literary text ranging from all sentences in some book or books containing a selected key word or words to an entire literary corpus. First however I should like to give a brief account of some earlier work by Prof. Myron Wish which has now become more or less classic. I know that there are problems with his interpretation of his data and that the program he used is now quite dated but its essential simplicity has helped others to get at least some grasp of the underlying aims of my current research and I repeat it without apology.

In 1968 Professor Myron Wish of the University of Wisconsin asked 18 of his students to rank 12 countries of the world in terms of their "overall perceived similarity," this without indicating any basis for these judgements since, as he noted, one of his aims was to discover the bases which they actually used. After averaging their scores and analyzing the result with the KYST multi-dimensional scaling program, he noted that the pro-Western and pro-Communist countries were in one set of diagonally opposite



McGill's Second Kierkegaard Collection

Figure 10. Three different accounts of Socrates.

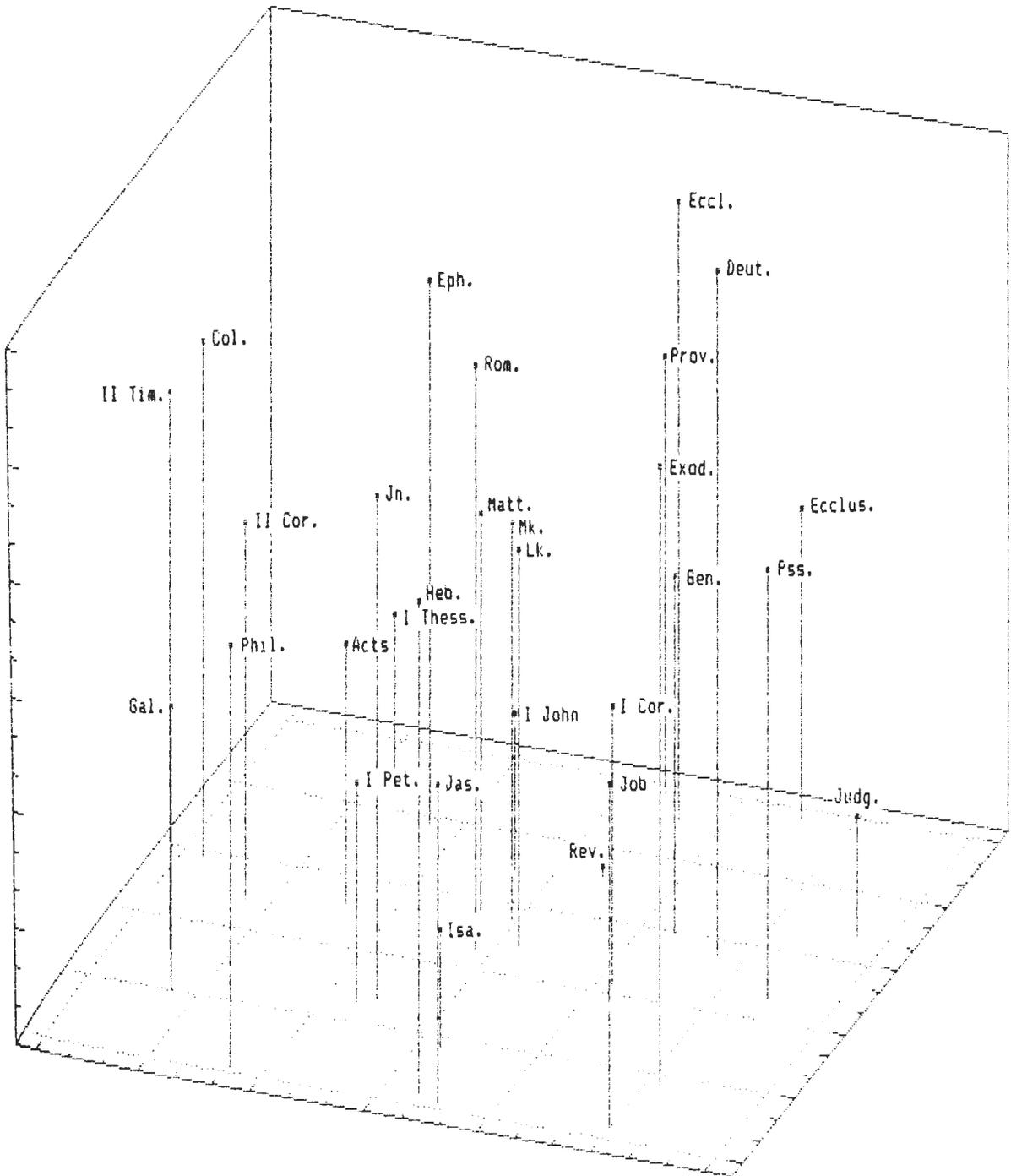


Figure 11. Kierkegaard's perception of the Bible.

quadrants, that the developed and underdeveloped ones were in the other and that these two dimensions together "explained" almost all of the variation in the original data matrix. Given these facts he concluded that these dimensions should be called "political alignment" and "economic development" and that his students had rated these countries primarily in light of these two features. Of course this does not necessarily mean that they consciously used these features but it clearly suggests that they made their actual judgements of similarity primarily in these terms. Details aside, I believe that Wish was entirely right and that his conclusions provide a radical insight into his students' collective perception of the relations of these countries. The aim of all my work in recent years has been to provide a similarly radical insight into Kierkegaard's works and the mind behind them and to show others how to do the same with the author in whom they are particularly interested.

This is a crucial point and should perhaps be explained more fully. I claim that Wish's dimensions provide a radical insight into his students' perception of the relations of these countries because they appear to be the bases upon which their particular judgements of similarity were actually made or, put another way, because they constitute the underlying structure of their collective perception of these relations. When we know the names of these dimensions we can understand or make sense of the two-dimensional array of their perception, explain why a particular country is here rather than there and, for example, even predict where these same students would place the next five countries within this space. My aim was to produce programs which would provide similarly radical insight into literary and philosophical texts. Of course, such texts are much more complex than 18 student replies to a questionnaire but real understanding is the same in both cases and fortunately we now have a much more powerful program capable of plotting many points in dimensions the number of which can equal that of actual text.

An aberrant frequency (hereafter, *abfreq*) word is one which an author has used in a par-

ticular book, for example, with a frequency statistically significantly greater than his normal one for that word.<sup>30</sup> The identification of such words is done by program and involves the following steps. First, divide the number of word-tokens in the book by the number of tokens in some appropriate control corpus and multiply by the frequency of that word in that corpus in order to determine its expected frequency in the book in question. Second, if that expected frequency is 20 or less, keep the word only if its observed frequency meets the criteria spelled out in our original account and shown here as Figure 12.<sup>31</sup> Third, subtract the expected frequency of this word from its observed and divide the result by the square root of the expected to determine its number of standard deviations from the corpus norm, generally known as its Z score. Note that a Z score of 1.96 means that this word can be expected to occur with this or some greater frequency in a book of this length by this author only five in 100 times and one of 2.58 that it could be expected to occur with such frequency only once in 100 times. Put another way, a word with a Z score of 3.29 can be expected to show such a frequency in only one in a 1,000 texts of this length by the author in question.

Our *abfreq* program generates two separate lists or files, one ordering these words according to their Z scores and the other alphabetically to facilitate comparison of related word forms.<sup>32</sup> In order to save space we show only the "top" and "bottom" of the Z score list for FB compared with Kierkegaard's *Samlede Værker* as a whole in Figure 13.<sup>33</sup> Note that this list has been expanded to provide translations and that even its last word shows a Z score of 9.77. This means that it could be expected to occur six or more times in only approximately one of every 10,000,000 texts of this size by Kierkegaard. Obviously these six uses of this word in this book represents such an extreme departure from his normal rate that it must have been deliberate and intentional. This must mean that its presence and use must be explained by any scholar claiming to understand or interpret this work and, equally clearly, that its real explanation must shed important light on this work.

| $\hat{\lambda} >$ and $w_i <$ |    | $\hat{\lambda} <$ and $w_i >$ |    |      |    |
|-------------------------------|----|-------------------------------|----|------|----|
| 3.7                           | 1  | 0.025                         | 0  | 9.1  | 15 |
| 5.6                           | 2  | 0.2                           | 1  | 9.9  | 16 |
| 7.3                           | 3  | 0.6                           | 2  | 10.5 | 17 |
| 8.8                           | 4  | 1.1                           | 3  | 11.5 | 18 |
| 10.3                          | 5  | 1.6                           | 4  | 12.0 | 19 |
| 12.0                          | 6  | 2.2                           | 5  | 13.0 | 20 |
| 13.0                          | 7  | 2.8                           | 6  | 14.0 | 21 |
| 14.5                          | 8  | 3.4                           | 7  | 14.5 | 22 |
| 16.0                          | 9  | 4.1                           | 8  | 15.3 | 23 |
| 17.0                          | 10 | 4.8                           | 9  | 16.0 | 24 |
| 18.5                          | 11 | 5.5                           | 10 | 17.0 | 25 |
| 20.0                          | 12 | 6.2                           | 11 | 18.0 | 26 |
|                               |    | 6.9                           | 12 | 18.5 | 27 |
|                               |    | 7.6                           | 13 | 19.3 | 28 |
|                               |    | 8.4                           | 14 | 20.0 | 29 |

Figure 12. Criteria for abfreq words with expected frequency = < 20.

Figure 14 shows one abfreq word from another and quite different point of view. Specifically, it shows the Z score of the intimate or familiar second person pronoun *Du* in each of Kierkegaard's works.<sup>34</sup> Thus we see at a glance that it has Z scores of approximately 55 in EE2, 33 in CT and 28 in TS and *minus* Z scores of approximately 27 in AE, 24 in SV and 21 in BI, the last three reminding us that low frequencies can be just as important and revealing as high ones. In fact this figure shows that we can and should think of all these statistically significant plotted Z scores as attributes or vectors of these particular books. Indeed, we can think of all the abfreq

words of a book as together constituting its distinctive attributes.

The usual next step in our method is to have the computer generate a data matrix or contingency table showing the frequency of each abfreq word in each chapter or section of the book in question similar to that shown as Figure 15. Note that this table uses my own abbreviations of the chapter titles of the recent Hong and Hong translation of FB, modified and explained in Appendix B, that it contains only some of the top 60 words from the previous abfreq list and that I have substituted their usual alphabetical order for one which

McGill's Second Kierkegaard Collection

| Word           | Translation/Use | Z score | Freq. |
|----------------|-----------------|---------|-------|
| Abraham        | Abraham         | 97.26   | 231   |
| Isaak          | Isaac           | 69.21   | 111   |
| Almene         | universal       | 44.57   | 104   |
| Agnete         | Agnes           | 42.16   | 40    |
| Sara           | Sarah           | 38.81   | 35    |
| Havmanden      | the merman      | 36.36   | 29    |
| Abrahams       | Abraham's       | 35.86   | 35    |
| Ridder         | knight          | 32.86   | 50    |
| Helt           | hero            | 30.56   | 71    |
| Kniven         | the knife       | 26.43   | 18    |
| Agamemnon      | Agamemnon       | 25.36   | 14    |
| Troens         | of faith        | 24.88   | 76    |
| tragiske       | tragic (=e)     | 24.63   | 39    |
| Iphigenia      | Iphigenia       | 23.39   | 12    |
| Troen          | faith           | 22.52   | 109   |
| Morija         | Moriah          | 21.30   | 10    |
| Absurde        | absurd          | 21.30   | 31    |
| Enkelte        | individual      | 20.84   | 103   |
| Morijabjerget  | Mount Moriah    | 20.26   | 9     |
| Resignationens | of resignation  | 19.71   | 16    |
|                |                 |         |       |
| Søn            | son             | 10.45   | 20    |
| Forjættelsen   | the promise     | 10.12   | 6     |
| Sønnen         | the son         | 10.07   | 11    |
| Faust          | Faust           | 9.98    | 16    |
| drog           | drew            | 9.97    | 9     |
| Fortællingen   | narrative       | 9.95    | 13    |
| Bjerget        | mountain        | 9.84    | 10    |
| Store          | the great       | 9.79    | 21    |
| intellektuel   | intellectual    | 9.77    | 6     |

Figure 13. Abbreviated abfreq list for FB, Z score version.

groups those having roughly similar profiles. This is a simple but crucial notion which we now explain briefly.

As the reader can readily determine for himself, Figure 15 shows that 31.7% of all cases of *universal* occur in P1, 40.4% in P2 and 27.9% in P3; that 4.9% of all cases of *individual* occur in PEX, 28.4% in P1, 37.3% in P2 and 29.4% in P3; and that 94.7% of all cases of *resignation* occur in PEX and 5.3% in P2. These, together with their equally important non-occurrences in all other chapters of this book, define or constitute the *profiles* of these words and are illustrated in Figure 16. The profiles of each of the chapters are calculated in precisely the same way, i.e. using the frequency of each of its abfreq words

expressed as a percentage of the frequencies of all its abfreq words. These profiles are the basis upon which all forms of correspondence analysis assign all word and chapter points to their proper position in multi-dimensional space and hence the key to the understanding of its operation.

Before proceeding to explain correspondence analysis we pause to note that our change program uses these same data matrices in a slightly different way. Briefly, it is a "moving window" chi-square test which begins by calculating the percentage of all occurrences of all abfreq words accounted for by the occurrences of each such word in the first chapter as against all remaining ones, in the first two against the remainder, etc., expresses these

McGill's Second Kierkegaard Collection

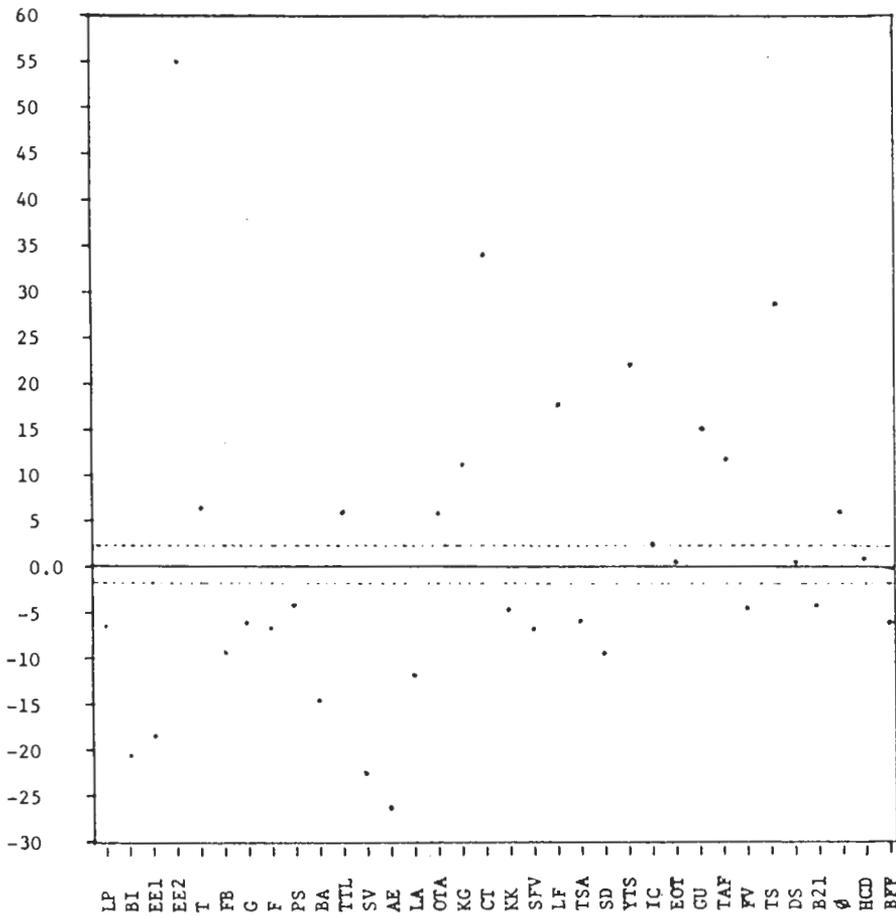


Figure 14. Z scores of  $D_u$  in all works of the authorship.

differences as a chi-square for each word, sums these chi-squares and identifies the point showing the largest total chi-square as the first or most important change. It then repeats this operation on both sequences of chapters on either side of this first break and continues thus until it has identified the number of changes requested or finds that none of the remainder are statistically significant. In fact the output shows the total chi-square value for every cut and both the actual and percentage contribution to this total of every word whose frequency shows a statistically significant change. For example, and as the full contents

of Figure 15 would show, *universal* accounts for 0.0% of all occurrences of all abfreq words in the first four chapters of FB but for 8.7% of all these occurrences in its last four chapters, a difference which represents a chi-square of 67.21. Of course such changes signal important shifts in the focus, subject and sometimes even tone of a book. No doubt a careful reader or, yet more likely, someone examining this matrix closely might identify some of these changes for himself but it is doubtful that anyone could identify all such changes or describe them in such detail as this relatively simple program.

*McGill's Second Kierkegaard Collection*

|                | PRF | EXM | ELY | PEX | P1 | P2 | P3 | EPG |
|----------------|-----|-----|-----|-----|----|----|----|-----|
| Mount Moriah   | 0   | 8   | 7   | 2   | 1  | 0  | 1  | 0   |
| drew           | 0   | 5   | 4   | 0   | 0  | 0  | 0  | 0   |
| rode           | 0   | 5   | 0   | 2   | 0  | 0  | 0  | 0   |
| the knife      | 0   | 4   | 7   | 4   | 1  | 0  | 2  | 0   |
| Sarah          | 0   | 7   | 11  | 0   | 0  | 1  | 4  | 0   |
| the hero       | 0   | 0   | 8   | 2   | 3  | 0  | 9  | 0   |
| the hero's     | 0   | 0   | 2   | 0   | 1  | 1  | 2  | 0   |
| believed       | 0   | 0   | 13  | 10  | 1  | 1  | 2  | 0   |
| Abraham        | 0   | 21  | 48  | 59  | 29 | 19 | 55 | 0   |
| son            | 0   | 2   | 3   | 2   | 4  | 3  | 4  | 0   |
| Isaac          | 0   | 23  | 17  | 27  | 7  | 16 | 21 | 0   |
| Abraham's      | 0   | 9   | 9   | 6   | 5  | 1  | 5  | 0   |
| the son        | 0   | 2   | 1   | 3   | 4  | 1  | 0  | 0   |
| the distress   | 0   | 0   | 0   | 0   | 5  | 2  | 4  | 0   |
| the dread      | 0   | 0   | 0   | 0   | 3  | 1  | 2  | 0   |
| tears          | 0   | 0   | 0   | 3   | 8  | 0  | 8  | 0   |
| universal      | 0   | 0   | 0   | 0   | 33 | 42 | 29 | 0   |
| individual     | 0   | 0   | 0   | 5   | 29 | 38 | 30 | 1   |
|                |     |     |     |     |    |    |    |     |
| movement       | 1   | 0   | 0   | 32  | 0  | 4  | 20 | 0   |
| courage        | 0   | 0   | 0   | 14  | 1  | 9  | 15 | 0   |
| narrative      | 0   | 2   | 0   | 5   | 1  | 1  | 4  | 0   |
| of resignation | 0   | 0   | 0   | 12  | 0  | 0  | 4  | 0   |
| the movements  | 0   | 0   | 0   | 6   | 0  | 0  | 1  | 0   |
| the princess   | 0   | 0   | 0   | 10  | 0  | 0  | 0  | 0   |
| the knight     | 0   | 0   | 0   | 9   | 0  | 0  | 0  | 0   |
| resignation    | 0   | 0   | 0   | 18  | 0  | 1  | 0  | 0   |

Figure 15. Abbreviated and re-ordered data matrix for FB.

Correspondence analysis is a relatively new form of exploratory data analysis which has proven extremely well suited for use with personal computers.<sup>35</sup> It is a powerful and highly specialized dual form of multi-dimensional scaling the goal of which is to represent the information contained in a data matrix in a more readily intelligible graphic display of lower dimensionality. Briefly, it uses eigenvalues and eigenvectors to compare the profiles of all word points in terms of their degree of similarity and determine their precise location in one true multi-dimensional space. It then does the same with all chapter points to determine their location in another such space. Finally, it superimposes one of these spaces upon the other to form a new joint multi-dimensional space in which all word and chapter points are arrayed in their proper relation to each other.

As already indicated, the program locates both word and chapter points according to the similarity of their respective profiles. Thus it clusters points with similar profiles into sets or clouds and separates those with the most dissimilar profiles as widely as possible throughout the available space. In fact it identifies the two most distant clusters as the termini of the first dimension, the next most distant clusters as the termini of the second, etc. Thus we can think of the dimensions of the space of a work as lines passing through its centroid or point of origin and joining clouds of points having very different profiles. Note that, unlike many earlier forms of multi-dimensional scaling, correspondence analysis makes these dimensions coincident with the actual geometrical axes of its space.

Another great advantage of correspondence analysis over other forms of multi-dimensional scaling is that it provides as supplementary information the percentage contribution of each word and chapter point to the total inertia of each dimension and indicates the pole of that dimension to which it contributes. This enables the scholar who already knows or is prepared to study the text carefully to name these dimensions. Of course the names of these dimensions together describe the ordered space of the book.

Note that it follows from the last two paragraphs that the dimensions of a space are identified by the program and are objective and real features of that space but that the names of these dimensions must be supplied by the researcher using the information provided by the program and his own preferably very considerable knowledge of the text. Briefly, the latter requires careful and prolonged reflection upon the two sets of chief contributors in order to discover what is common to these sets (the name of the dimension) and what distinguishes them (the names of the two poles). A difficult but very rewarding task!<sup>36</sup>

We have already explained how correspondence analysis orders points in space and we now offer a preliminary account of how such an array can help us to understand a book. Of course everyone assumes that a book is a physical object consisting of printed pages but this is a gross confusion of the means with the message. In fact, the real book is much more like a vision in the mind of the author which he is attempting to convey to the mind of the reader through the awkward and often distorting medium of strings of linear text conforming at least generally to the local language and grammar. That vision is much more like an ordered space and can be represented more perspicuously and accurately as an array of its key words in multi-dimensional space, whether metaphorically in the mind of its author or literally as the profiles of the words and chapters (or other divisions) in the printed text. As already explained, correspondence analysis enables one to model these visions or, as I prefer, to re-create the space of the book or text in question. Briefly, it allows the user to name its dimensions,

describe its space, identify its most closely related words and concepts and, as we shall see, even construct its own multi-dimensional concordance.

It is perhaps worth noting one other reason why correspondence analysis results are so important. Books are generally read for what they say but the most important part of any book is precisely what it does *not* say. There are many reasons why such parts are regularly omitted. The author may be unconsciously or even consciously unable to write what he most wants to say. The point he wishes to make may be simply unsayable within his historical context. He may regard certain points as so obvious and self-evident that he would never think of asserting them. The last is particularly important because nothing is so difficult to see or so important to notice as the obvious. The approach we have been describing underscores the obvious and repeatedly calls attention to what the text assumes but does not explicitly say.

#### INTERLUDE: FROM DATA TO GRAPHICS

The preceding account attempts to explain how our software converts merely numeric data into meaningful graphics but may be too complicated and technical for some readers. We therefore pause briefly to describe a very simple example showing how one can move from mere word frequencies to a graphic display of the relations of the books in which they occur. The steps are shown in successive columns of Figure 17 and in Figures 18 and 19. We begin with the now familiar example *Du*.

Column A of Figure 17 shows the absolute frequency of *Du* in each of Kierkegaard's works ordered chronologically. Note the wealth of information and absence of or at least difficulty in discerning any trend or pattern. Column B shows these same books again ordered chronologically but this time with relative rather than absolute frequencies thus permitting meaningful comparisons between books of very different sizes. For example, it shows that *Du* is approximately 29 times more frequent in the relatively small YTS than in the very large AE but the chronological order still makes it difficult to detect any trends. Column C overcomes this difficulty by order-

McGill's Second Kierkegaard Collection

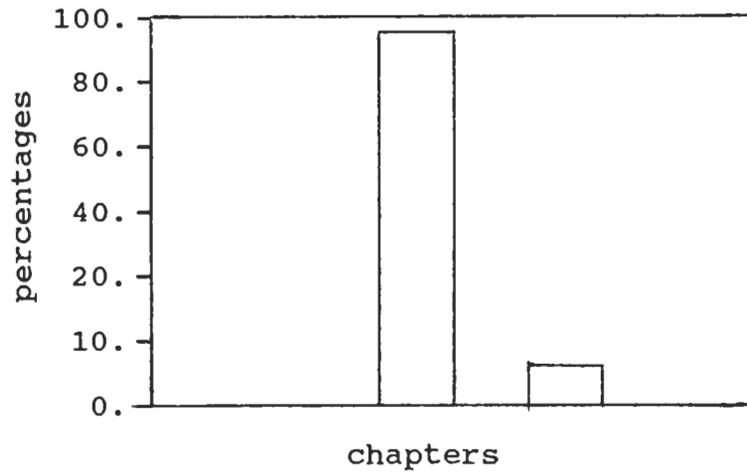
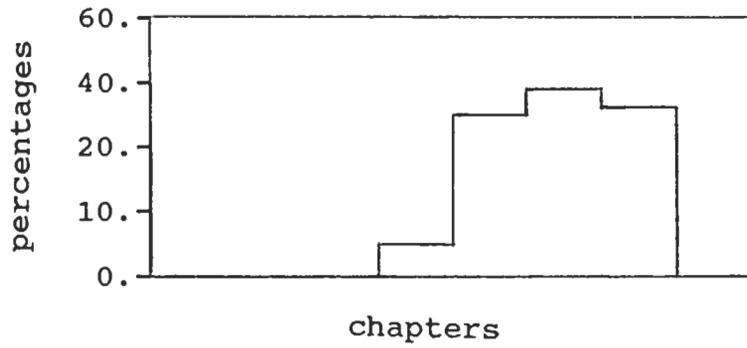
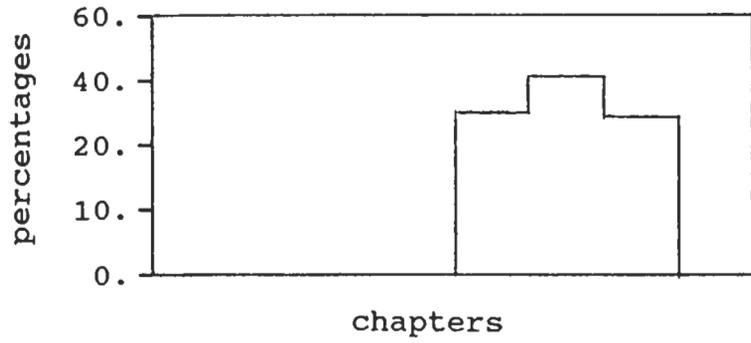


Figure 16. Profiles of *Almene*, *Enkelte* and *Resignation* in FB.

McGill's Second Kierkegaard Collection

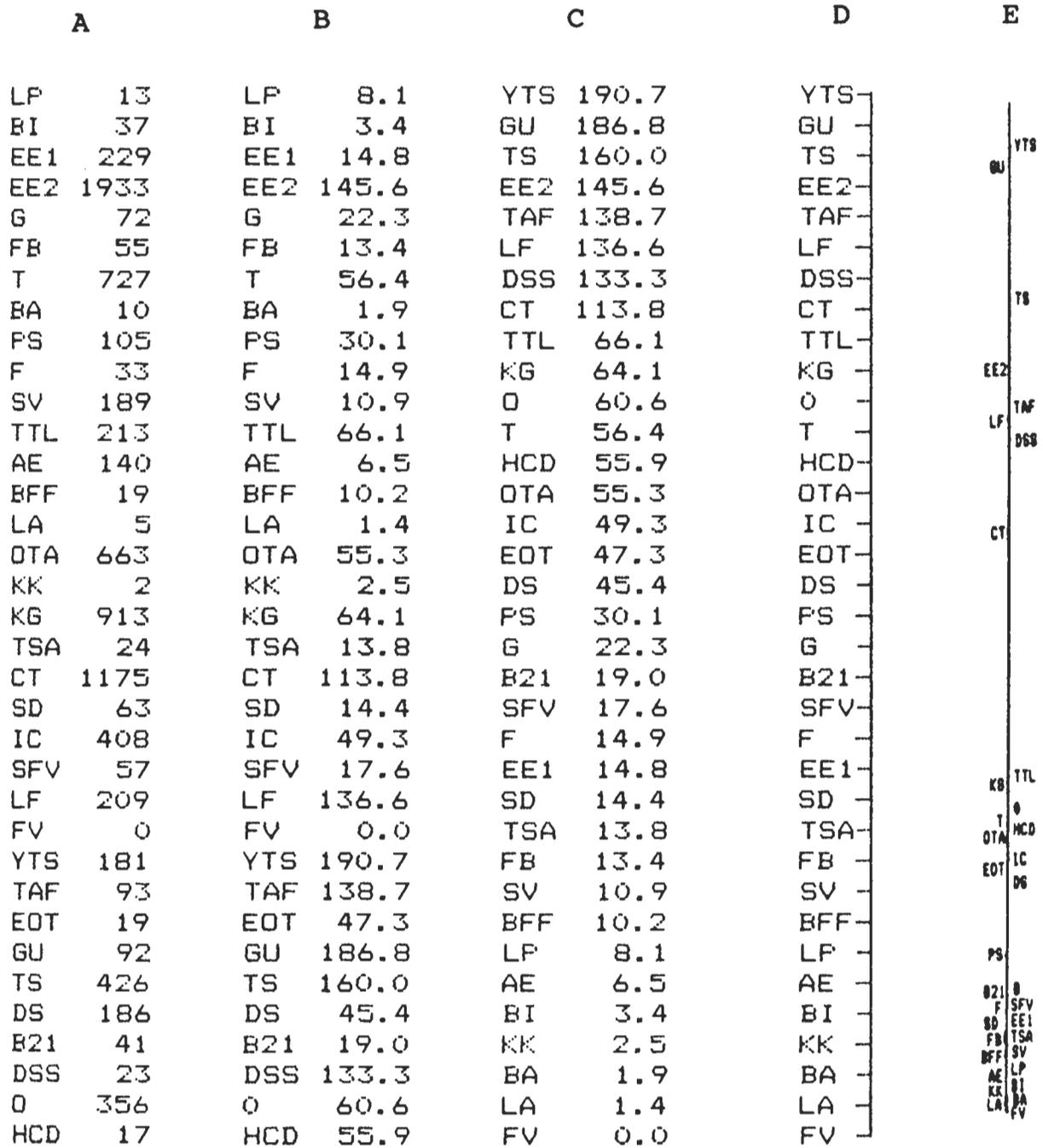


Figure 17. Comparison of various tabular and graphic displays.

ing these books according to the relative frequencies of this word. Note, for example, the large differences between the frequencies of DSS and CT and, especially, between CT and TTL. Note too the large number of works with frequency less than 10. Column D shows these books placed in this same order at regular intervals on a simple line graph and therefore actually contains less information than the preceding one. By contrast, column E plots these books using the relative frequencies of *Du* as coordinates and shows in a clear and graphic way the relation of these books to one another with respect to this particular but important attribute or property.

Figure 18 plots these same books using the relative frequencies of *Du* as the *y* and those of *man* (one, people) as the *x* coordinates. This plot shows the relations of these books to one another with respect to both these attributes simultaneously and hence is much richer than its one-dimensional counterpart. Note, for example, that the addition of *man* has separated some devotional works (T, TTL, OTA and KG) from some previously grouped "attack" works (IC, DS, HCD and O). Figure 19 adds a third dimension and arrays these points along it using the relative frequencies of *jeg* (I) as the *z* coordinates. This completes this process and produces a richer array which shows some interesting clusters (e.g. EOT, TSA and TAF, FV and SFV, and TS and DSS) but which in general does not reflect the actual relations of these works. But our real purpose in building up these two arrays is to stress that the very great difference between them and the kind produced by correspondence analysis, a difference frequently overlooked and even denied by those who fail to understand the latter. In the present two cases the points are plotted according to their relevant coordinates and the dimensions in question simply represent different relative frequencies for that variable. In short, these are simple two- and three-dimensional plots of the kind with which we are all familiar. In the case of correspondence analysis, however, the position and hence the coordinate of every point on every dimension is a complex function of the relations between its profile and that of all other points in the set. Thus correspondence analysis does not simply place

points along pre-defined dimensions but analyses the profiles of all the points to determine the dimensions of the space and the coordinates of each point on each of its dimensions. In this case the user does not tell the program where to place the points but rather provides a matrix giving their profiles and allows it to determine their space and their position within it. This is something quite different and is why, to use our own phrase, it can be used to recreate the space of a book.

Correspondence analysis allows the user to create multi-dimensional models of concepts, books and corpora as well as multi-dimensional concordances, etc. These appear to be the most important and promising results described in this study and some examples will be provided shortly. First, however, we briefly report some abfreq and change point studies which, separately and together, can tell us a great deal about the history of a concept, book or corpus.

## THE USES OF ABFREQ WORDS

"Kierkegaard's Attack upon Christendom: Its Lexical History"<sup>37</sup> was one of my first and simplest abfreq-based studies and, as its title was intended to suggest, traces the history within the authorship of the words most characteristic of that attack. Briefly, it identified the first 60 abfreq words in B21, DSS, O and HCD, the four late works generally accepted as constituting this attack, and traced back through the authorship to the work in which these were first generally used in an "aggressive" sense. Our investigation showed that 45 of these words had already been so used much earlier, many as early as AE, that nine of the remaining 15 reflect the deliberately journalistic and provocative character of this attack and that another three are connected with its immediate occasion. In short, the essential vocabulary of the attack was already "on stream" or "in place" (i.e. had already been "shared" with his readers) well before the final attack itself was actually launched. Of course this means that the attack is much more closely connected with the earlier authorship than generally supposed and, I believe, suggests a new and promising interpretation of AE. Naturally it would be a very simple matter

McGill's Second Kierkegaard Collection

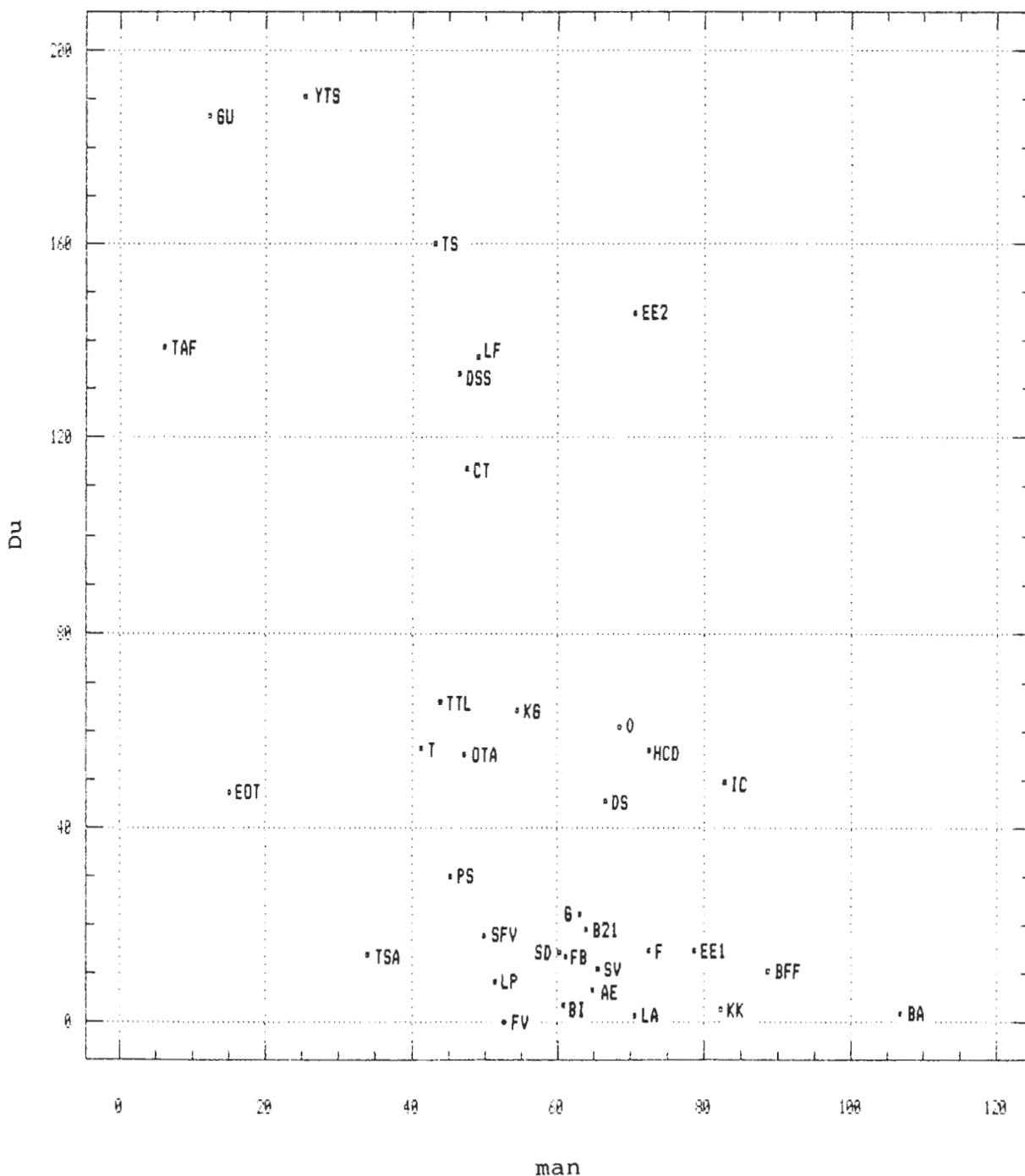


Figure 18. Books using *man* and *Du* as coordinates.

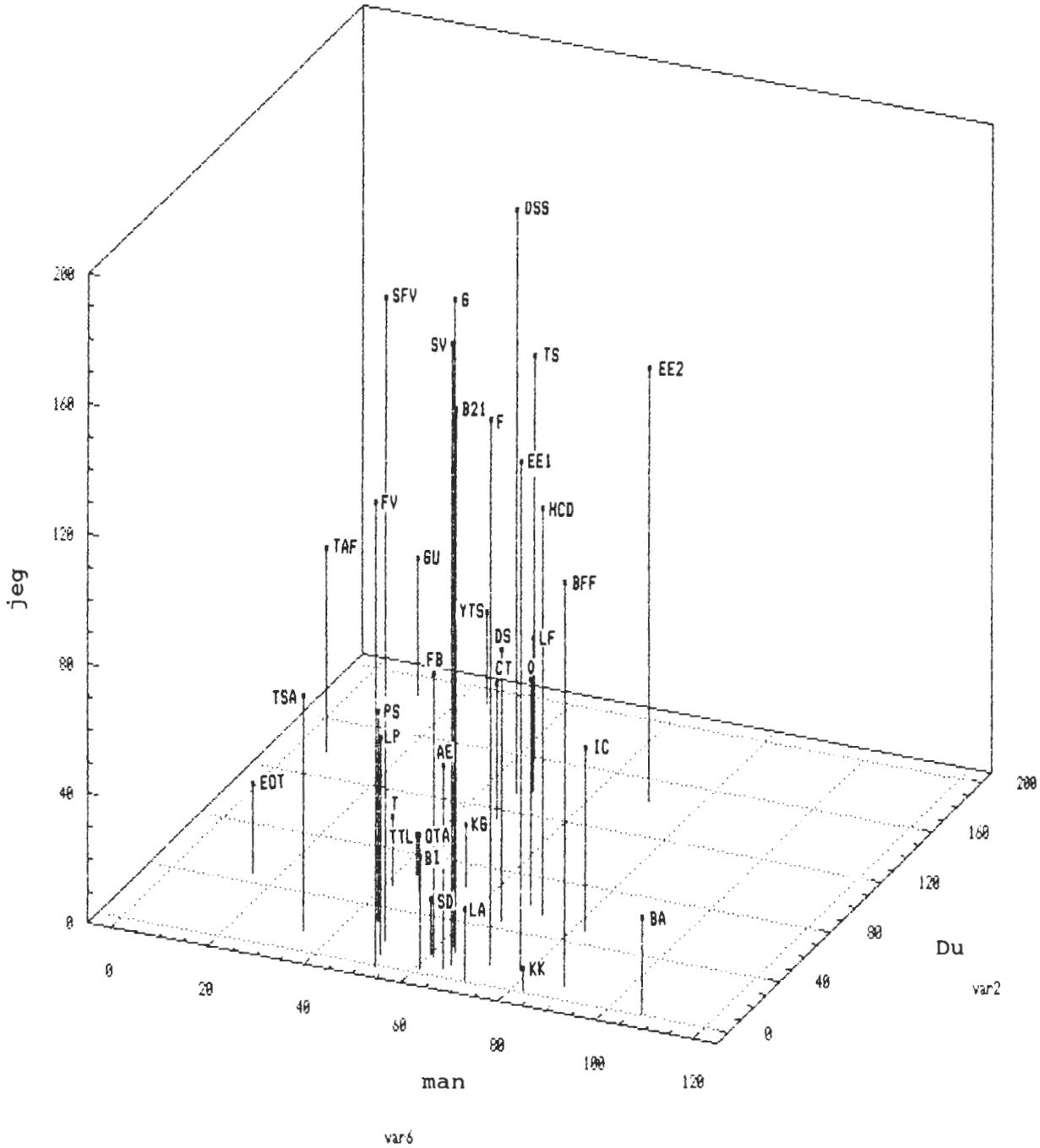


Figure 19. Books using *man*, *Du* and *jeg* as coordinates.

to adapt this method to trace the history and development of a theme in any other text or corpus.

A more recent and much more complex study identifies the abfreq words associated with God and with Christ in the 19 books of the authorship in which one or, in most cases, both figure significantly. Briefly, having identified these words in these books treated cumulatively, i.e. in the first book, in the first two books, in the first three books, etc., it then documents the "transfer" of words previously associated with one to the other. The results show that the accounts of God and Christ generally become more and more alike with the passage of time. This convergence is clearly illustrated and its rate shown in Figure 20. Of course this method can also be used to do studies of divergence or convergence of two or more characters, families, traditions, themes, etc. across a book or corpus.

A current abfreq study compares Kierkegaard's accounts of the lilies and the birds which, as in the gospels, are widely assumed to have similar characteristics and roles. Figure 21 contains two relevant short excerpts from the output of our COMPARE program which compares any two similar files, in this case the abfreq files containing the 60 words most closely associated with *Lilien* (the lily) and *Fuglen* (the bird) in Kierkegaard's four accounts of these figures. This program shows that these two lists have only 3 or 2.564% of their words in common,<sup>38</sup> namely, *anxious one*, *anxiety* and *be silent*. It is obvious then that while Kierkegaard revered both these "teachers" he nevertheless conceived and presented them in quite different ways and that an investigation of the words unique to each will enable us to tease out these differences. It is also clear that these programs can alert their user to many important and obvious differences which traditional scholarship frequently misses.

## CHANGE POINT STUDIES

One of the change point studies we have done but not published uses as input the data matrix showing the frequencies of the first 60 words in the eight chapters of FB, part of which matrix has been shown as Figure 15.

As already explained, the method identifies the most important changes in these data, reports its chi-square value and that of individual words together with an indication of whether their proportional frequency is increasing or decreasing. The output of this program is necessarily extensive and, for the moment at least, not as transparent as one might wish. We have therefore devised a simple graphic which shows the five statistically significant changes in these data and which is shown as Figure 22. This graphic says that the first or most important change or shift occurred *between* the first four and the last four chapters in this book and that the chi-square value for this change is the very high 713.82. Similarly, it shows that the second and next most important change is *between* P1 and P2 on the one hand and P3 and EPG on the other and that the chi-square for this change is 345.92. Etc., etc. Note further that the original output also lists the words making the largest contributions to each of these changes together with their respective chi-squares and an indication of whether they are becoming a larger or smaller proportion of the whole. Thus it reports that *Almene* (universal), *Enkelte* (individual) and *Helt* (hero) show a proportional increase after the first change or cut, that the chi-squares of their changes are 67.21, 51.12 and 36.54 and that these represent 9.4, 7.2 and 5.1 percent, respectively, of the total chi-square of this change. Similarly, it reports that *Morijabjerget* (Mount Moriah) and *Troen* (the faith) show a proportional decrease after this cut, that the chi-squares of these changes are 35.27 and 26.44 and that these scores represent 4.9 and 3.7 percent of the total for this change. These results provide a complete and useful summary of the many changes in the proportional frequencies of key words in a text and the user who studies them carefully will find them helpful and revealing.

Another such study documented the changes in the frequencies of the abfreq words associated with *det Gode* (the Good) in the nine works in which this word occurs most frequently. Our simple graphic summary of these results is shown in Figure 23. Note that the first and second change occur immediately before and after OTA which contains Kierkegaard's only sustained treatment of this

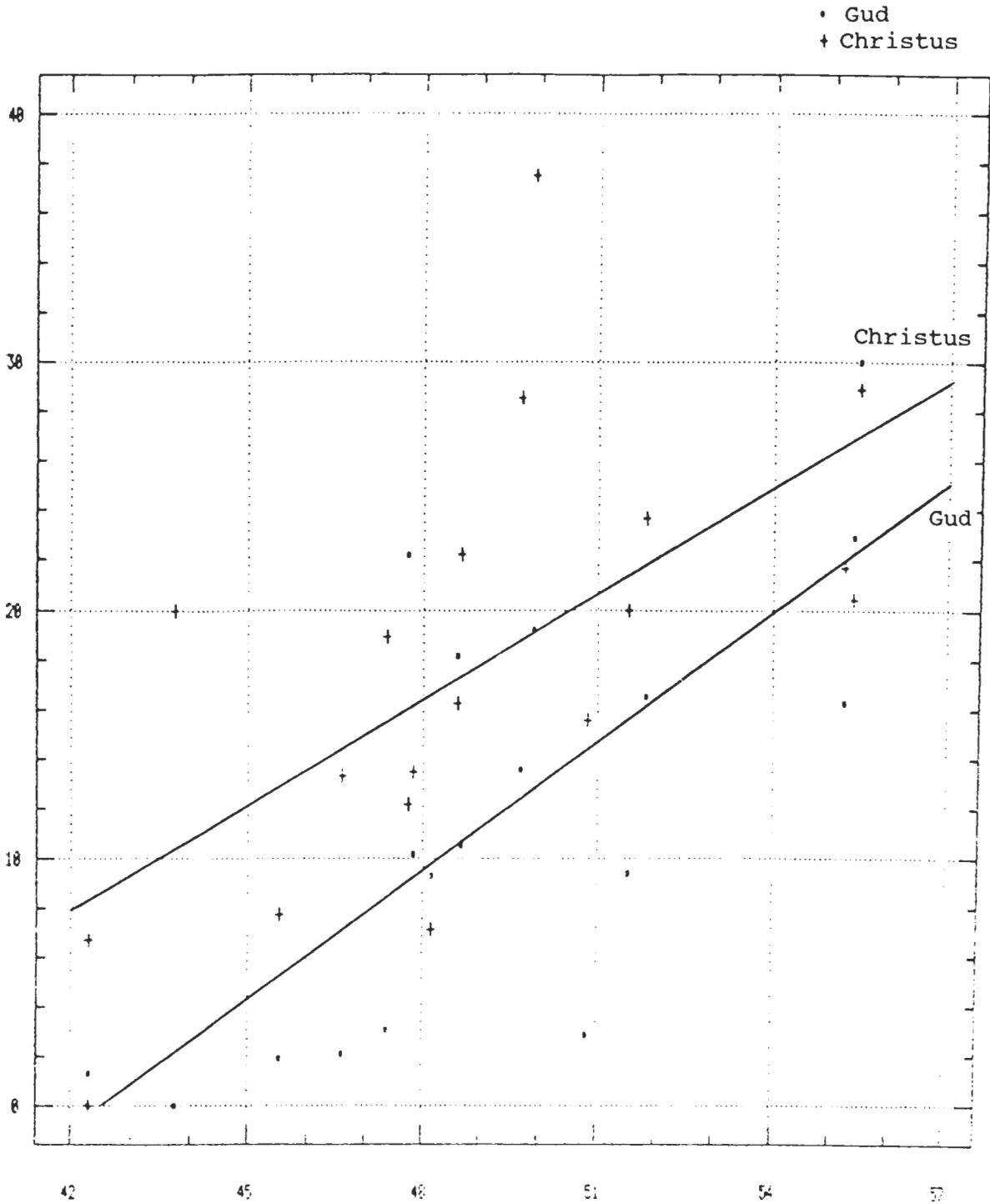


Figure 20. Convergence of *Gud* and *Christus* across authorship.

McGill's Second Kierkegaard Collection

|             |           |           |
|-------------|-----------|-----------|
|             |           | Afgrunden |
| agte        |           | Armodens  |
|             | Bekymrede |           |
| bekymrede   |           |           |
| bekymres    |           |           |
|             | Bekymring |           |
| Bekymringer |           |           |
| Betragter   |           |           |
|             |           | Brød      |
|             |           | Fuglen    |
| Lilien      |           |           |
| syer        |           |           |
| Syerske     |           | synger    |
| tale        |           |           |
|             | tier      |           |
|             |           | trælle    |
| ubetinget   |           |           |
| uheldigt    |           |           |
| Undergang   |           |           |
|             |           | underveis |
|             |           | uvidende  |

57 words unique to 'lil160.abf'.

57 words unique to 'fug60.abf'.

3 words appear in both files.

Percent common = 2.564 %

Index/Concordance = 0.026

Figure 21. Abfreq words peculiar and common to *Lilien* and *Fuglen*.

theme, that the dashed line indicates that none of the remaining changes are significant at the 0.05 level and that there are more and larger changes before OTA than after it. Of course the printout also shows the contribution of each word to each change, many of which are themselves statistically significant even when the overall changes themselves are not. It thus provides the user with many, many clues as

he seeks to understand and document more fully all of the various changes in the development of this or some other concept.

Another combined change point and classical chi-square study entitled *Dating Kierkegaard's Battles with Fate*<sup>39</sup> compared the respective frequencies of all forms of *Skjebne* (Fate), *Forsyn* (Providence) and

*Styrelse* (Governance) across Kierkegaard's *Samlede Værker* and *Papirer*, the latter as reported in Cappelørn's *Index*. The results of this study for the former of these sources are shown in a somewhat different graphic in Figure 24. Note that *Skjebne* accounts for a quite large percentage of all occurrences of these words in the writings from around 1840 but for a much smaller one in those from the end of 1851. By contrast, *Styrelse* accounts for only a very small percentage of all occurrences of these three words in the early writings but for a much larger percentage in the later ones. In fact this graph can be read as depicting a very important aspect of Kierkegaard's life and thought. Naturally, of course, similar studies can be done on other authors and their works on themes in which one is interested.

Finally, our MAP program shows the relative frequencies of a word or family of words in both tabular and graphic form. Those for the seven main forms of *Paradox* are shown in Figure 25 and Figure 26. Note that the former shows the works chronologically in three columns and uses two asterisks to indicate that these forms are most frequent in PS, that they have a relative frequency of 26.08 and so are to be given the highest shading level (8) in the map. The latter, Figure 26, arrays the works along a time line representing the years of the authorship with the acknowledged works above and the pseudonymous below. The base line of a work represents its period of composition, its total area its actual size and its shading level the frequency of these forms in this work. Note that most of the acknowledged works show no occurrences of any of these forms and that BI has only 2 and LA and O only one each. In short, they are virtually absent from the acknowledged works of the authorship.

## CORRESPONDENCE ANALYSIS

The studies just described are obviously important but we think that those employing correspondence analysis have greater promise and potential. Briefly, they can recreate the space of a text, identify its clusters, name its dimensions and plot its argument; all this for any kind of text including all sentences containing some selected word or words in one

or more books, a book or set of books or even an entire corpus. We now briefly describe some of these studies beginning with two which used our own CORRESP program which is part of the collection and which accepts only a smaller input matrix.

One of my own first correspondence analysis studies used a matrix showing the frequency of Kierkegaard's 50 most common nouns in each of his 35 books to array these books in two dimensions and thus provide a very simple but reasonably accurate picture of their relation to each other and to the whole. To the best of my knowledge this array has not been previously published but I know that it hangs on the wall of several Kierkegaard scholars and I show it here as Figure 27 for any who may find it helpful in understanding the relations of the individual books to one another and to the authorship as a whole. Note that I have grouped these works and named the groups the aesthetic, the religious, the attack and the "meta" works, the last of course being the works about the works. Of all these placings the only one which seems to me at all problematic is that of AE within the "meta" works and it is clear even from this array that it might perhaps have been included instead within the aesthetic. However it contains an extended discussion of the earlier books, a final and crucial statement about their place in the authorship, was later described by Kierkegaard as its "turning point" and, at least on my reading, is among other things a preview of the attack. Of course it is possible that, given more data or a third dimension, this work might clearly separate itself from the other three in this group but with this proviso I believe that this array can be accepted as a simple, memorable and substantially accurate account of the various relations of these books. Of course I also believe that this simple approach would yield at least equally reliable results for the vast majority of other and less complex authors.

Several years ago I was invited to attempt to resolve the question whether Kierkegaard could have been the author of "Litterært Qvægsølv" ("Literary Quicksilver," hereafter LQ). For this study I used the same correspondence analysis program and the frequen-

McGill's Second Kierkegaard Collection

| Cut | PRF   | EXM   | ELY   | PEX   | P1    | P2    | P3    | EPG   | chi-sq. |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|---------|
| 1.  | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- | 713.82  |
| 2.  | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- | 345.92  |
| 3.  | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- | 311.30  |
| 4.  | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- | 206.92  |
| 5.  | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- | 135.54  |

Figure 22. Summary of changes in FB.

| Cut | EE2   | T     | BA    | AE    | OTA   | KG    | CT    | SD    | IC    | chi-sq. |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|---------|
| 1   | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- | 526.30  |
| 2   | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- | 401.54  |
| 3   | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- | 151.36  |
| 4   | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- | 104.12  |
| 5   | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- | 75.19   |
| 6   | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- | 70.00   |

Figure 23. Summary of changes in accounts of *det Gode*.

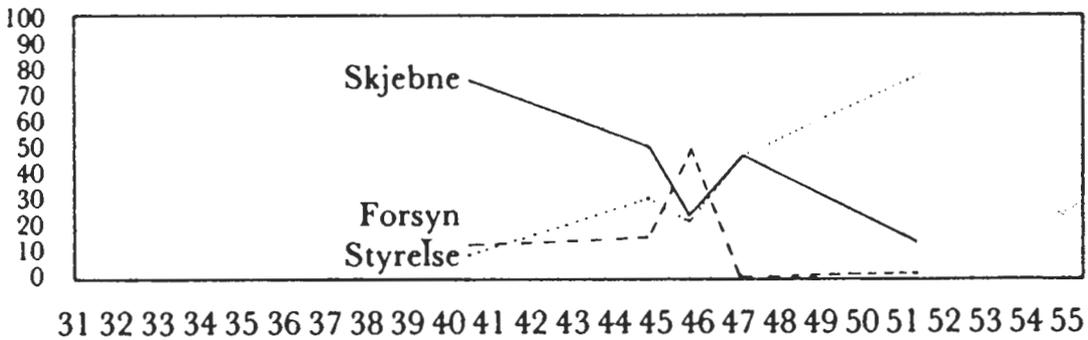


Figure 24. Changes in frequencies of *Skjebne*, *Forsyn* and *Styrelse*.

McGill's Second Kierkegaard Collection

|        |           |      |           |      |          |
|--------|-----------|------|-----------|------|----------|
| LF:    | 0.00 (0)  | AE:  | 10.13 (4) | FV:  | 0.00 (0) |
| BI:    | 0.19 (1)  | BFF: | 0.00 (0)  | YTS: | 0.00 (0) |
| EE1:   | 1.16 (1)  | LA:  | 0.28 (1)  | TAF: | 0.00 (0) |
| EE2:   | 0.15 (1)  | OTA: | 0.00 (0)  | EOT: | 0.00 (0) |
| G:     | 0.00 (0)  | KK:  | 0.00 (0)  | GU:  | 0.00 (0) |
| FB:    | 18.83 (6) | KG:  | 0.00 (0)  | TS:  | 0.00 (0) |
| T:     | 0.00 (0)  | TSA: | 4.61 (2)  | DS:  | 0.00 (0) |
| BA:    | 0.00 (0)  | CT:  | 0.00 (0)  | B21: | 0.00 (0) |
| ** FS: | 26.08 (8) | SD:  | 3.67 (2)  | DSS: | 0.00 (0) |
| F:     | 0.00 (0)  | IC:  | 0.97 (1)  | O:   | 0.17 (1) |
| SV:    | 0.06 (1)  | SFV: | 0.00 (0)  | HCD: | 0.00 (0) |
| TTL:   | 0.00 (0)  | LF:  | 0.00 (0)  |      |          |

Figure 25. Frequencies of main forms of *Paradox*.

cies of the comma, semi-colon, colon, period, exclamation and question mark which earlier tests showed to be his most sensitive punctuation discriminators. The correspondence analysis plot for LQ and nine of Kierkegaard's similar short writings from the same period are shown in Figure 28 and that for LQ and all of his 35 books in Figure 29. Note that in both cases LQ stands clearly apart from all the other texts and is distinguished mainly by its higher number of exclamation and question marks. This, together with the results of chi-square tests for the nine similar pieces, prompted me to conclude that it is extremely unlikely that this work could have been written by the author of these pieces or, indeed, of Kierkegaard's published writings. The need for and importance of authorship determination methods is obvious; this one has the advantage of providing both a probable answer and an explanation of its basis.

More recently I have used Professor Michael Greenacre's SimCA program to analyze a matrix containing the frequencies of the 250 most common nouns and names in 34 books<sup>40</sup> of the corpus. The purpose of this study, published as "Mapping the Dimensions of a Literary Corpus,"<sup>41</sup> was to name the first eight dimensions of the authorship and thus shed light on its underlying oppositions and structure. A very brief and somewhat updated graphic summary of these results is provided

in Figure 30. Note that these eight dimensions account for 63.7% of the information in the input matrix and together presumably describe the most important of its many famous dialectical emphases. For example, the first line says that the strongest contrast or opposition in the authorship is that between the many early aesthetic and the middle period religious works and that this dimension accounts for 29.60% of the information in the original matrix. The fifth says that it presents the two main forms of the God relation as faith and despair and that this accounts for 9.02% of this information.

Yet more recently I have used this same program in an attempt to gain some insight into Kierkegaard's much celebrated but, I believe, very perplexing FB. In this study I analysed a matrix showing the frequencies of the first 124 abfreq words in each of its eight chapters using the same abbreviations as before. That study is not yet finished but I believe that I have succeeded in naming the first seven dimensions of this work. These are shown in a similarly simple form in Figure 31. The first line suggests (correctly and obviously) that the main division within this work is that between the accounts of Abraham's deed and those of the problems arising, especially that of his silence. The fifth line (and dimension) underscores the crucial fact that this work contains two quite different

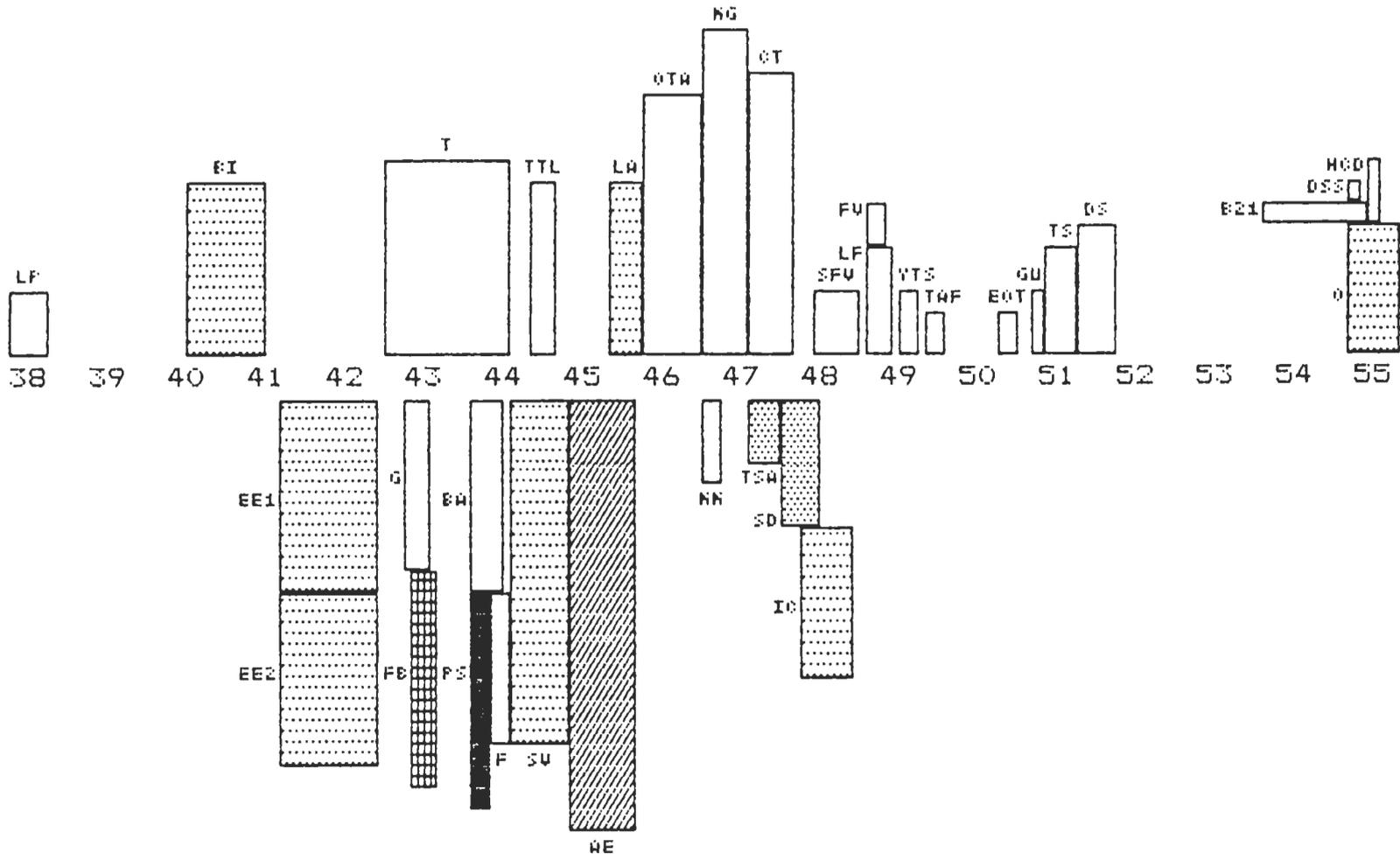


Figure 26. Main forms of *Paradox* in works of corpus.

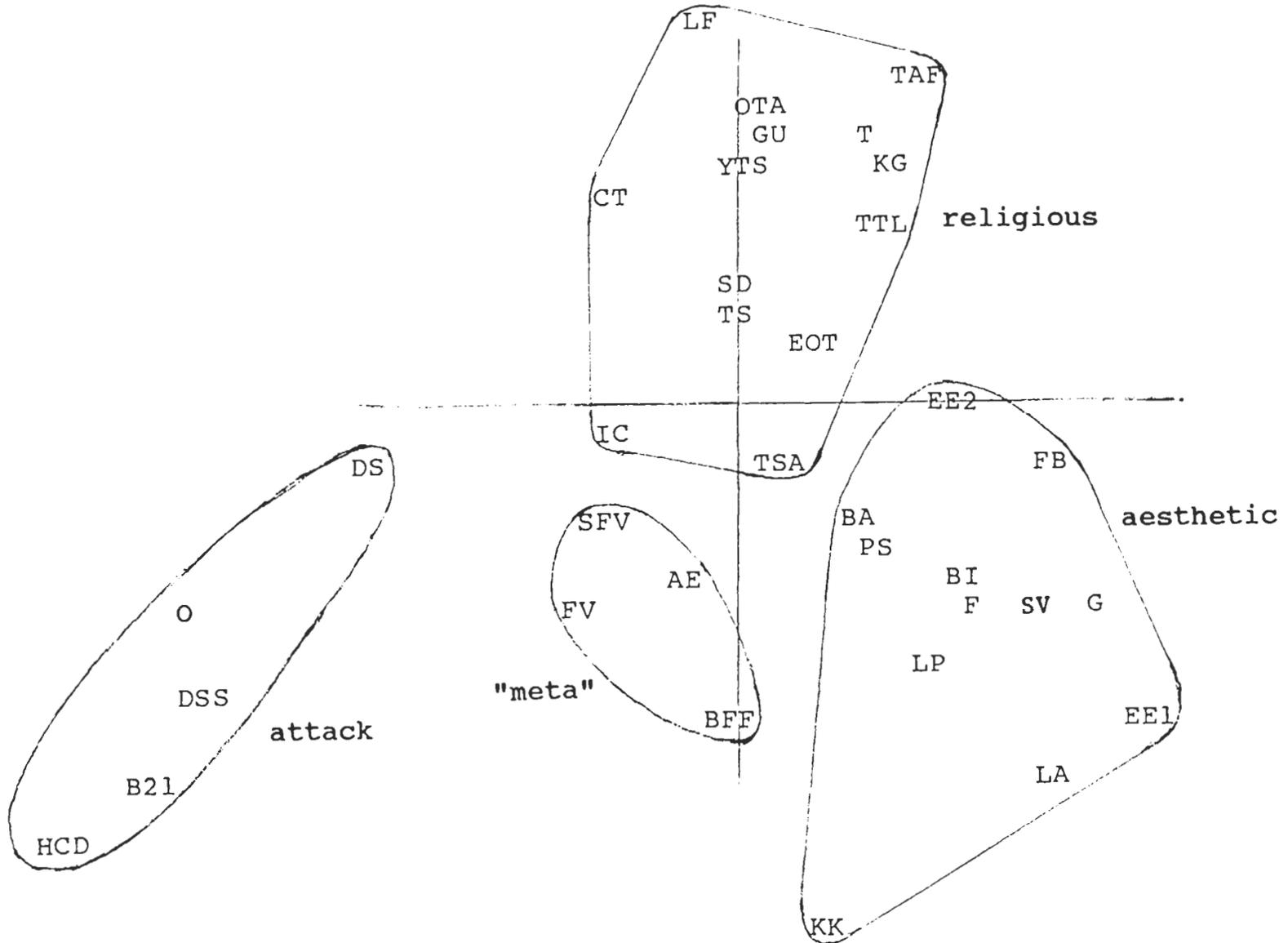


Figure 27. Array of works using frequencies of 50 most common nouns.

McGill's Second Kierkegaard Collection

|                      |              |
|----------------------|--------------|
| C2                   | ;            |
| C6<br>C1<br>C4       | LQ           |
| C7<br>C9<br>C3<br>C8 | C5<br>?<br>! |
| :                    |              |

Figure 28. Array of LQ and nine similar works using punctuation.

|                 |                        |
|-----------------|------------------------|
| :               | FV                     |
| YTS             | KK<br>SFV LA           |
| HCD B21<br>ØOTA | AE BI                  |
| EOT IC          | SD                     |
| TAF DS LF       | LPTSA BFF EE2 BA       |
| TS              | TTL<br>;CTKG FSVFB EE1 |
| GU              | T PS                   |
| !               | ?                      |
| LQ              | G                      |

Figure 29. Array of LQ and works of corpus using punctuation.

McGill's Second Kierkegaard Collection

| # +                            | Dimension                        | - %   |
|--------------------------------|----------------------------------|-------|
| 1. early aesthetic.....        | MAIN THEMES.....middle religious | 12.85 |
| 2. aesth. & relig.....         | THE WHOLE AUTHORSHIP.....attack  | 9.81  |
| 3. love the neighbour.....     | ENDS OF MAN.....worship God      | 9.14  |
| 4. of society.....             | PATHOLOGY.....of self            | 7.58  |
| 5. faith.....                  | THE GOD RELATION.....despair     | 6.99  |
| 6. as status.....              | THE INDIVIDUAL.....as task       | 6.59  |
| 7. God.....                    | OBJECTS OF DEVOTION.....the Good | 5.78  |
| 8. the individual in time..... | TIME.....the God in time         | 4.93  |

Figure 30. The first eight dimensions of Kierkegaard's authorship based on frequencies of 250 most common nouns and names.

| # -                           | Dimension                                 | + %   |
|-------------------------------|---|-------|
| 1. problems, esp. silence.... | MAIN THEMES.....Abraham's deed            | 29.60 |
| 2. ethical.....               | RIVAL PERSPECTIVES.....religious          | 23.93 |
| 3. his silence.....           | JUSTIFICATION OF ABRAHAM.....his deed     | 21.13 |
| 4. ethical.....               | JUSTIFICATION OF HIS DEED.....religious   | 11.01 |
| 5. unbeliever.....            | ACCOUNTS OF ABRAHAM.....believer          | 9.02  |
| 6. resignation.....           | SILENTIO'S DISTINCTION.....faith          | 3.59  |
| 7. "beyond faith!".....       | RIVAL VIEWS OF FAITH...."highest passion" | 1.71  |

Figure 31. The first seven dimensions of FB using 124 abfreq words.

accounts of Abraham, the first as unbeliever and the second as believer.

Figure 32 shows the space of FB in terms of its first three dimensions and traces the course of its "argument" through this space. Note that it begins with the Preface which appears to point to the justification of Abraham's deed from a religious perspective, moves to the Exordium where Abraham is presented as an unbeliever and from an ethical point of view, then to the Eulogy where he is presented as a believer and apparently from a less strictly ethical perspective. It then moves to the Preliminary Expectoration in which Johannes *de silentio* introduces a notion of faith which includes but goes beyond that of mere resignation. It next moves to the discussion of the problems raised by Abraham's deed (Problem 1 and Problem 2) and, espec-

ally, his silence (Problem 3), all considered from a mainly Hegelian ethical perspective, rather than Silentio's own properly religious perspective. Finally, it moves back to the Epilogue thus ending virtually where it began. Of course none of us can visualize more than three dimensions simultaneously but note that any three of these seven can be combined in such an array and the argument of the book traced through its space thus re-created.

In fact this same study had another and perhaps even more important outcome. Long and detailed study of its results and careful comparison with the text eventually led me to see that this program had actually given me what can only be called a multi-dimensional concordance to this work. Of course it has been widely assumed for more than twenty years that computers were ideally suited to produce

concordances just because they reproduce fragments of the original linear text. In fact I now see that this is precisely why they have not met and cannot meet our expectations. As I have already argued, books are not strings of linear text but rather visions in the minds of their author, visions which are best understood as arrays of their key words in multi-dimensional space. Correspondence analysis enables us to recover that vision or recreate that space and so provides a multi-dimensional concordance giving the context of any word in three quite new and different senses of that word. These I describe as the spatial, the overall and the role/sense context. The first is the *position* of the word in the named and ordered space of the work, the second the desired number of words closest to the word in the full space of the work and the third the desired number of words associated with that word in one of its various roles or senses and lying within a "cone" of the space radiating from its origin, adjacent to the vector of that word and running out to the edge of the space.

These new and different kinds of context can be illustrated with examples from FB the basic story of which is familiar to all. "Abraham" lies in the "Abraham's deed" (as opposed to the "problems arising") area of the work, has as its overall context the words son, should, father, God, Isaac, soul, God's, deed, was tried, the best, sacrifice (n.), etc. and as its role/sense context the words drew, expectation, Mount Moriah, rode, the mother, was tempted, Sarah, the knife, the Lord, etc. I believe that no one with a thorough knowledge of this text will deny the accuracy or relevance of any of these contexts. Consider, for example, that the accounts of Abraham as doer of the deed typically report that he was tempted (*fristet*) by the Lord (*Herren*) while those dealing with the problems arising say instead that he was tried (*forsøges*) by God (*Gud*). Note, too, that the words comprising the overall context (i.e. those connected with Abraham as he is presented in the book as a whole) are quite distinct from those comprising the role/sense one (i.e. those connected with him as "doer of the deed") and that, in principle at least, it is possible to provide all these contexts for any word in the original

text. Finally, note that this concordance helps the reader distinguish the different roles or senses of a word rather than simply confronting him with all its instances as they happen to occur in the text. In fact it represents a very great advance over the traditional linear concordance and in a better and different world would shortly replace it.

Of course this proposal to replace linear concordances with multi-dimensional ones represents a re-thinking of an earlier and important part of this entire project. In fact I now see that we should go back even further and return to the pseudonyms project with which it all began. The reason is obvious. Correspondence analysis can array the various pseudonyms in multi-dimensional space and thus provide a much more subtle, detailed and accurate account of their relation to Kierkegaard and to one another than the essentially one-dimensional one we produced in 1967. Of course this means going back to the beginning but there is nothing wrong with that. As an older Danish friend has always said, and as our own models of Kierkegaard's works have constantly shown, "all good books end where they begin."

#### PRINTED PAGE OR COMPUTER SCREEN?

Much of the preceding has described various programs and studies designed to help one discover and identify structures, patterns and connections within text and we now take the further and apparently outrageous step of suggesting that the computer screen is in many respects actually better suited for serious scholarly reading than the familiar printed page or, at the very least, that it has certain very important advantages. Of course many will dismiss this as unthinkable heresy but this is perhaps because they simply associate reading with books and fail to distinguish between reading for pleasure and reading for serious scholarly purposes. In fact, alleged eye problems, radiation hazards and hard chairs notwithstanding, the computer screen offers the scholar a number of important advantages which we detail in terms of the present collection and as briefly as possible.

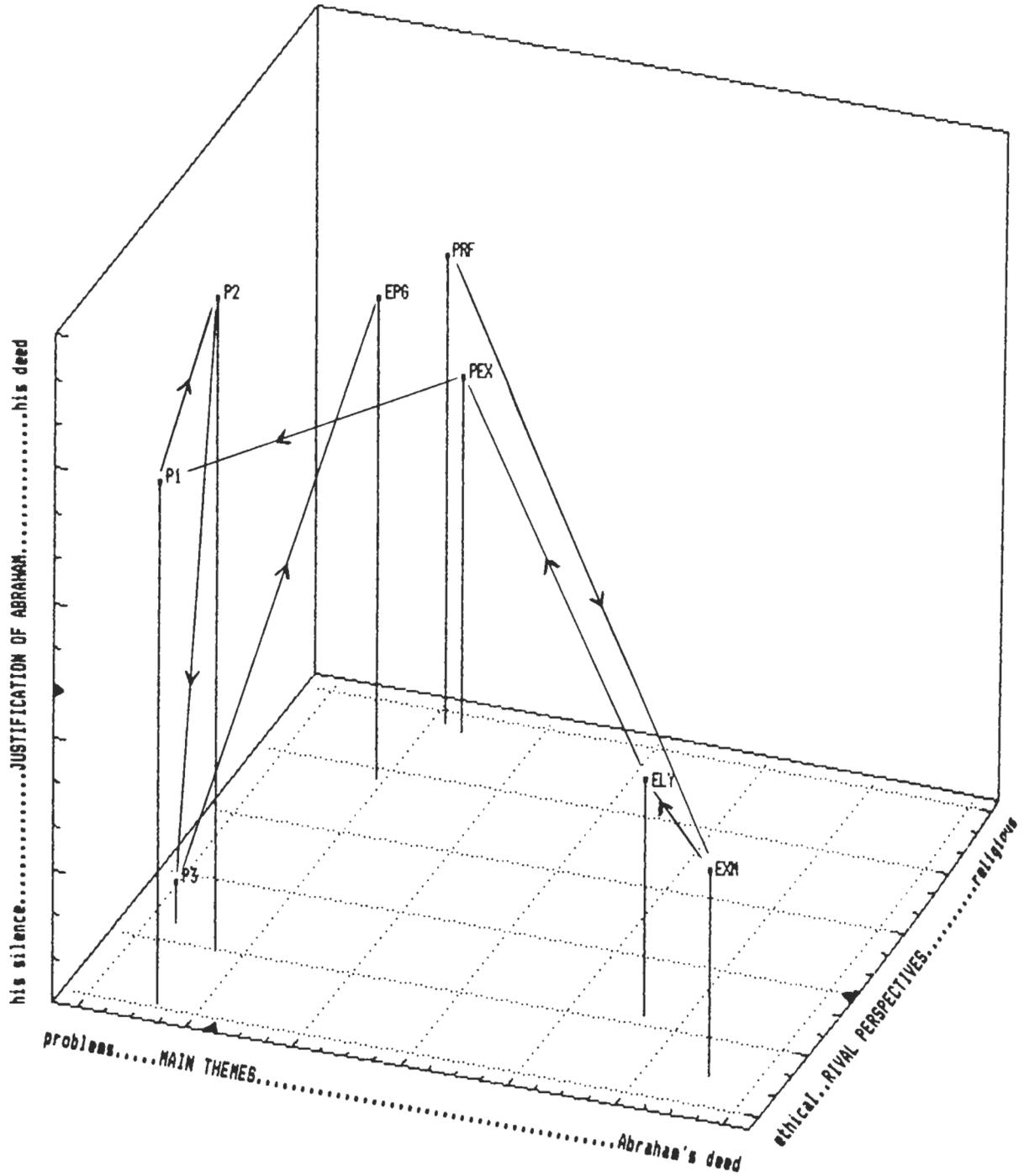


Figure 32. The "argument" of FB plotted in three dimensions.

The scholar using our display and search program can readily locate every occurrence of a given word in one or all texts of the third or "base" edition and instantly obtain on request its precise location in any or all of the "secondary" editions included in this collection. For example, he can display KG on the screen, type *Bedrag* (deception) on the status line and examine at his own pace every occurrence of this word in its complete context and, if he wishes, immediately consult another edition or translation of his choice. This is immensely faster than the traditional visual search and much less prone to distraction by other interesting passages.

Of course he can also locate and display every occurrence of a given concept as distinct from its various word forms. For example, he can type the string *edrag* and see in succession and at his pleasure every occurrence of every form of the concept deception (deception, deceptions, the deceptions, of deception, to deceive, deceive, deceived, deceiving, deceptively, etc.). Obviously this is very important for the scholar attempting to sense or taste the full range of the concept he is investigating.

Note that both the searches described above allow the user to obtain the location of the current instance in another edition or translation of his choice. This provides simultaneous and immediate access to both the original text and one's most trusted translation, saves much time, effort and confusion, rightly acknowledges the primacy of the original text and frees the user from undue dependence upon the translator.<sup>42</sup>

Of course the scholar can also request the program to display any specified page and line in the text. In fact it will do this in a fraction of the time it takes one to find the same location in the printed text and at the same time provide any or all "secondary" references as a bonus.

Note that another program in the collection allows the user to extract the line, three lines, sentence or paragraph containing some specified word or words, save these extracts in a file showing their location in the "base" and some other edition and print out this file for subsequent study, notation and comment.

Further, another simple program allows the user to define any block of text and send it to a word processing program or editor for inclusion in a scholarly paper or book together with the secondary reference of his choice. Depending upon the amount of text involved, the saving in time and effort can be substantial. Of course one may xerox printed texts but the results are rarely suitable for pasting into a paper.

Finally we note that a number of electronic texts are already significantly cheaper than their printed counterparts and that their prices will almost certainly continue to decline. Indeed the difference is already so great that, all the old prejudices notwithstanding, it may soon do more to persuade scholars to use electronic text than all the considerations cited in this account.

## CONCLUSION

All scholars know how to use traditional collections but this one is very different and consists of two separate though connected parts. We conclude by commenting briefly upon their present and possible uses both here at McGill and in the wider world of scholarship.

Imagine a scholar attempting to solve the question of Kierkegaard's use of *Spring* (leap) or of his even more difficult notion of *Anfægtelse* (spiritual trial [?]). One having access only to the Malantschuk collection would almost certainly begin with the works containing earlier uses of these terms with which Kierkegaard was known to be familiar and by which he was probably influenced. He would probably also consult Molbech's *Dansk Ordbog*, the 1819 Danish translation of the New Testament, etc. By contrast, one having access also to our second collection would, in the first instance at least, probably use it in one or more of the ways we have described in an attempt to understand more clearly how Kierkegaard actually used these words in his own works. Briefly, the Malantschuk collection gives insight into the historical background while the second collection provides new ways of discovering connections within the text. Hence, though very different, the

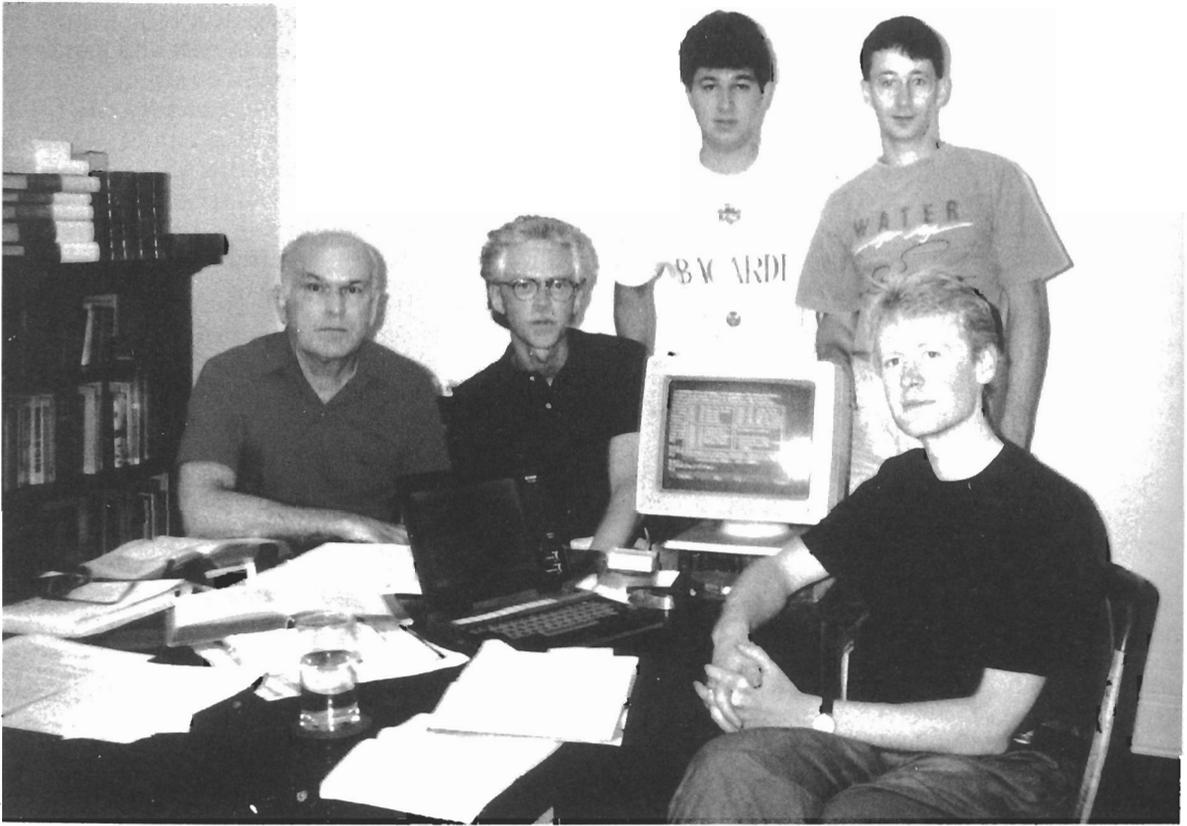


Figure 33. The Canadian-Danish Kierkegaard critical edition team, Montreal, July, 1991.  
(L to R): Alastair McKinnon, Niels Jørgen Cappelørn, Nikos Krivossidis,  
Stéphane Hogue and Joakim Garff

two collections complement each other perfectly which, as already noted, appears to have been the chief reason Malantschuk decided that his own collection should come to McGill.

The texts of traditional collections are invariably from the past and are rightly valued because they are historic. By contrast, ours is merely a corrected version of the Danish third edition published in 1962-64 and still currently available. However unless I am greatly mistaken it too will prove to be a truly historic text. It is already used throughout Canada and the United States and in every continent except Africa and is now so widely distributed that one or more copies should survive almost

any conceivable catastrophe. Of course individual copies can "go down" or be lost in a "crash" but they can be copied easily in virtually no time and at very little expense. It is already being used as the basis of the current joint Canadian-Danish project to create new computer and printed critical editions of both Kierkegaard's *Samlede Værker* and *Papirer* based upon his life-time editions and manuscripts. Given the rapidly increasing use of computer and related technology, it seems likely that this will be the main form in which this text will survive into and be used in the future. Of course it is a record of the past but at the moment its particular value and interest

would seem to be that it comes from and points to the future.

The reader is reminded that the software now included in this collection can be used only with the Kierkegaard or other similarly formatted text but that a version for use with free format text can be made available. In fact, Kierkegaard computer studies are already in relatively good shape at McGill and my main concern in recent years, and certainly my chief purpose in providing this account, has been to encourage my colleagues in the humanities to venture into this field, mainly by providing tools for their use. Of course computer studies of text continue to be demanding and it is still true that faint heart never won fair lady. However, compared with the past, they are now relatively easy. Depending upon the quality of the original, text can now be scanned with around 96% accuracy and computer proof-reading programs can raise this to 99.9%, punctuation being the chief remaining problem. A proven set of formatting conventions have been established and others are possible. Software is available and I am of course prepared to provide assistance with its use. I would however make one point which is so important that it must be mentioned separately.

This account has consisted mainly of brief descriptions of the results of some of my own studies because I could not think of any other or better way to describe this software to colleagues in the humanities but I want to stress as strongly as possible that it was not designed primarily for any such specific studies. In fact it consists of approximately 30 separate programs most of which are basic tools designed to be put together in a variety of ways and intended primarily to allow and indeed encourage the user to indulge and follow his imagination; in fact, this is the chief reason I have always resisted pressure to combine them into a pretty menu-driven package. My own studies have now been recognized and, though no doubt applicable to other authors, do not need to be repeated. Of course this software will always be here at McGill but it will be more than an historical curiosity only if others find new and different ways of using it. I therefore encourage my colleagues to take

up the challenge; after all it will be their use of these tools which, in the long run, will determine the worth and importance of this aspect of the collection.

Finally, a word about the relation of the university to the past and the future. Humanists of course have a systematic bias in favour of the past which is why they always think of collections as consisting of old documents. Of course the past is important and those who do not learn from it are condemned to repeat it, a truism which has a darker side which many prefer to overlook. But we humanists also need to learn to think more seriously about and plan for the future. We need to ask if there are not better ways of doing the important work we have been given and, if there are such ways, to learn to use them. I recall a brief but magic moment which happened only recently. After a long and serious discussion, the editor of Denmark's oldest and largest publishing house interrupted the formalities of our parting to say how interesting it had been to talk with someone who thought "not of next Thursday but of the next century." As we say, "That made my day."

\* \* \* \* \*

APPENDIX A

Title codes of works used in this account.

|     |                                      |                                      |
|-----|--------------------------------------|--------------------------------------|
| LP  | Af en endnu Levendes Papirer         | [From the Papers of One...]          |
| BI  | Om Begrebet Ironi                    | The Concept of Irony                 |
| EE1 | Enten – Eller. Første halvbind       | Either/Or, vol. 1                    |
| EE2 | Enten – Eller. Andet halvbind        | Either/Or, vol. 2                    |
| G   | Gjentagelsen                         | Repetition                           |
| FB  | Frygt og Bæven                       | Fear and Trembling                   |
| T   | Atten opbyggelige Taler              | Eighteen Edifying Discourses         |
| BA  | Begrebet Angest                      | The Concept of Dread                 |
| PS  | Philosophiske Smuler                 | Philosophical Fragments              |
| F   | Forord                               | [Prefaces]                           |
| SV  | Stadier paa Livets Vei               | Stages on Life's Way                 |
| TTL | Tre Taler ved tænkte Leiligheder     | Thoughts on Crucial Situations ...   |
| AE  | Afsluttende ... Efterskrift          | Concluding Unscientific Postscript   |
| BFF | Bladartikler, ..."Forfatterskabet"   | [Articles about the Authorship]      |
| LA  | En literair Anmeldelse               | Two Ages                             |
| OTA | Opbyggelige Taler ... Aand           | Purity of Heart, Gospel of Suffering |
| KK  | Krisen og en Krise ... Liv           | Crisis in the Life of an Actress     |
| KG  | Kjerlighedens Gjerninger             | Works of Love                        |
| TSA | Tvende ... Smaa-Afhandlinger         | Two Minor ... Discourses             |
| CT  | Christelige Taler                    | Christian Discourses                 |
| SD  | Sygdommen til Døden                  | The Sickness Unto Death              |
| IC  | Indøvelse i Christendom              | Training in Christianity             |
| SFV | Synspunktet ... Forfatter-Virksomhed | The Point of View ... an Author      |
| LF  | Lilien paa Marken og Fuglen ...      | "The Lilies of the Field and ... "   |
| FV  | Om min Forfatter-Virksomhed          | On My Work as an Author              |
| YTS | "Ypperstepræsten" ... "Synderinden"  | "The High Priest'... '... Sinner'"   |
| TAF | To Taler ved Altergangen ...         | Two Discourses at the Communion .    |
| EOT | En opbyggelig Taler                  | "An Edifying Discourse"              |
| GU  | Guds Uforanderlighed                 | God's Unchangeableness               |
| TS  | Til Selvprøvelse, Samtiden anbefalet | For Self-Examination                 |
| DS  | Dømmer selv                          | Judge for Yourself!                  |
| B21 | Bladartikler 1854-55 I-XXI           | Newspaper articles, 1854-5           |
| DSS | Dette skal siges; ...                | "This must be said ..."              |
| O   | Øieblikket, nr. 1-10                 | The Instant, nos. 1-10               |
| HCD | Hvad Christus dømmer ...             | "What Christ thinks ..."             |

APPENDIX B

Notes

Chapter titles of FB, modifications and abbreviations used in this account

|                              |     |
|------------------------------|-----|
| 1. PREFACE                   |     |
| Preface                      | PRF |
| 2. EXORDIUM                  |     |
| Exordium                     | EXM |
| 3. EULOGY ON ABRAHAM         |     |
| Eulogy                       | ELY |
| 4. PRELIMINARY EXPECTORATION |     |
| Preliminary Expectoration    | PEX |
| 5. PROBLEMA I                |     |
| Problema 1                   | P1  |
| 6. PROBLEMA II               |     |
| Problema 2                   | P2  |
| 7. PROBLEMA III              |     |
| Problema 3                   | P3  |
| 8. EPILOGUE                  |     |
| Epilogue                     | EPG |

\* \* \* \* \*

1. In all honesty I must report that both the University of Copenhagen and St. Olaf College in Minnesota, the other two major centres of Kierkegaard research, also have book and computer collections. However this one was created at McGill and can truly be described as a McGill collection.

2. This collection, including the text and software, together with all the research summarized in this account, was made possible by continuing grants from the Canada Council, the Humanities Research Council of Canada and, at times of particular difficulty, the Faculty of Graduate Studies and Research of McGill University. The four volumes of *The Kierkegaard Indices*, also briefly described, were published with assistance from the Humanities Research Council of Canada. I take this opportunity to record my sincere thanks to these agencies and, particularly, to the late Dean Walter Hirschfeld and Dean Gordon MacLachlan and, for constant support during the early stages of this work, to the staff of the McGill Computing Centre, especially Professor David Thorpe, its former director.

3. This collection was donated to McGill University in March 1988 with the encouragement of the late Dean Richard Salisbury and is available to members of the University at the Faculty of Arts Computing Laboratory in the Leacock Building together with a manual for its use. Of course I should be glad to help anyone wishing advice about the use of any of this text or software. I also note that the Laboratory now has similar resources for the study of Wittgenstein.

4. This and the other title codes identifying Kierkegaard's various works are explained in Appendix A.

5. Kierkegaard's *Samlede Værker* has been published in three separate editions only the last of which is now widely available. Our computer version contains many corrections but generally follows the pagination and lineation of this edition which is therefore treated as our "base" printed text. I am happy

to take this opportunity to record my deep thanks to Gyldendalske Boghandel of Copenhagen for the gift of several copies of this edition over the many years of this work.

6. Note that the first six columns of this line actually read "00913M" and that the use of such "leading code" in computer text has many important advantages. In fact, our display program "hides" this information and shows it instead on the status line in order not to interfere with the reading of the text.

7. "Column 01" means that the cursor is presently in the first column of this line and "Replace" that this older program was then in the "Replace" mode, i.e. that it would permit corrections to the text. It has now been replaced by a more recent version which accepts changes only after the user has supplied the correct password.

8. This corresponds to the main divisions of a book and hence to the various levels of indentation found in a good table of contents to a complex work.

9. This important piece was added to AE and, on his instructions, originally printed on unnumbered pages.

10. "There is something inexplicably felicitous in the antithesis: Climacus – Anti-Climacus, I recognize so much of myself and my nature in it that if someone else had invented it I would believe that he had secretly observed my inner being. – ..." *Søren Kierkegaard's Journals and Papers*. Ed. Howard V. Hong and Edna H. Hong. (Bloomington: Indiana University Press, 1978): 6532.

11. Alastair McKinnon, *Kierkegaard's Critique of Rationalism* (University of Edinburgh, Ph.D. thesis, 1950), 53.

12. This error resulted from two decisions innocent in themselves but fatal when taken together. Briefly, we defined a "pair-word" as one found in any Kierkegaard sample but in only one pseudonymous one and at the same time chose two samples from Johannes Climacus, one from his PS and the other from his AE. These decisions, taken together, reduced the ratio for the author of AE to slightly less than that for the author of EE1.

13. The notable exceptions, of course, are those who, for their own reasons, deliberately restrict their attention entirely to the pseudonymous works, often implicitly identifying one or other of these with Kierkegaard.

14. Alastair McKinnon, "Kierkegaard's Pseudonyms: A New Hierarchy," *American Philosophical Quarterly*, 6, 2, (1969): 124.

15. The Kierkegaard Indices consist of the following four volumes: Alastair McKinnon, *Kierkegaard in Translation: en Traduction: in Übersetzung* (Leiden, Brill, 1970); Alastair McKinnon, *Konkordans til Kierkegaards Samlede Værker* (Leiden, Brill, 1971); Alastair McKinnon, *Index Verborum til Kierkegaards Samlede Værker* (Leiden, Brill, 1973) and Alastair McKinnon, *Computational Analysis of Kierkegaard's Samlede Værker* (Leiden, Brill, 1975).

16. These are slightly larger than the comparable figures reported for this corpus in volume IV; the differences are due almost entirely to the subsequent inclusion of all Greek text.

17. Henry Kučera and W. Nelson Francis, *Computational Analysis of Present-Day American English* (Providence, Brown University Press, 1970).

18. Alastair McKinnon, *Computational Analysis of Kierkegaard's Samlede Værker* II, column 1.

19. Alastair McKinnon, "Kierkegaard's Remarks on Philosophy," *Journal of the History of Philosophy* XI, 4 (1973): 513-522.

20. Alastair McKinnon, "Theological Focus in Kierkegaard's *Samlede Værker*: Some Basic Data," *Studies in Religion/Sciences Religieuses*, 4, (1974-75): 58-62.

21. This date represents the midpoint between the date of commencement and completion and has been calculated for every work on the basis of the information provided in Alastair McKinnon and Niels-Jørgen Cappelørn, "The Period of Composition of Kierkegaard's Published Works," *Kierkegaardiana*, (1974): 133-146.

22. The relative frequency of a word is its frequency in the unit in question divided by the

number of word-tokens in that unit, itself divided by the figure (e.g. 100, 1,000, etc.) yielding the most appropriate result. In this case the figure used was 100 but in retrospect it seems that a larger number would perhaps have been better.

23. Alastair McKinnon, "The Increase of Christian Terms in Kierkegaard's *Samlede Værker*," *Kierkegaardiana* (1974): 147-162.

24. By "neutral" I mean terms which in general at least reflect preoccupation with as distinct from advocacy of Christianity. These words are listed in the study in question.

25. Niels Jørgen Cappelørn and Alastair McKinnon, "Kierkegaard's Literary Production by Quarterly Rates," *Danske Studier* (1982): 21-34.

26. Alastair McKinnon, "A Method of Displaying Differences between Various Accounts of an Object," *revue CIRPHO review* (1974): 31-57. At the request of Alessandro Cortese a later version of this study was published as an introduction to such work for Italian scholars; see Alastair McKinnon, "Three Conceptions of Socrates in Kierkegaard's Writings," *Kierkegaard Oggi* (Milano, Vita e Pensiero, 1985): 27-39.

27. KYST is the acronym for the Kruskal-Young-Shepard-Torgerson multi-dimensional scaling program written by Dr. J. B. Kruskal, Bell Telephone Laboratories, Murray Hill, N.J. and Dr. F. W. Young, Psychometric Laboratory, University of North Carolina, Chapel Hill, N.C., assisted by Judith B. Seery, Bell Telephone Laboratories, Murray Hill, N.J.

28. Alastair McKinnon, "Kierkegaard's Perception of the Bible," *Kierkegaardiana* (1980): 132-147. This study was based upon information contained in Paul S. Morimoto and Paul Minear, *Kierkegaard and the Bible: An Index*," Princeton (1953).

29. I say more about this below in connection with the notion of the "space" of a text but meanwhile remark that our current understanding of intelligence appears to be as skewed as the Stanford-Binet test which, as Prof. Howard Gardner has been recently pointed out, is really only a measure of the

candidate's ability to move quickly from one item to another and does not touch any of at least seven other aspects of intelligence, one of the most important of which is the ability to conceive relations in spatial terms. Another sign of this same bias: not long ago a distinguished philosopher confessed to me that he could not read or understand even the simplest time vs. temperature chart. What have our philosophers and educators done to us? To what end or purpose? And for whose benefit?

30. Alastair McKinnon, "Aberrant Frequency Words: Their Identification and Uses," *Glottometrika* 2 (Bochum): 1980 108-124. G.L.M. Berry-Rogghe is generally credited as being the first to conceive and implement the basic idea and it is perhaps not to my credit that I was not aware that she had done so. The solution to the problem of words having an expected frequency = < 20 has been devised by David Sankoff to whom I am most grateful.

31. This table was constructed from a table of Poisson probabilities in a way which assures that the probability of rejecting the null hypothesis is about 5% evenly distributed between the two tails of the distribution.

32. Note that the package also includes a program allowing the user to edit the first of these files and that it saves an updated version of both if any changes are made.

33. Note that, though abbreviated, this list extends down to the sixtieth word, this because our own correspondence analysis and change point programs can accept matrices of only up to 36 rows and 60 columns. Later in this report we describe some more recent studies of this book based upon a matrix using its first 124 abfreq words.

34. Note that the dotted lines in this figure represents Z scores of 1.96 and hence that the frequency of *Du* in all works shown above or below these lines is significant at the conventional 0.05% level.

35. Given present attitudes to this approach it is perhaps worth noting at the outset that correspondence analysis is not merely a current French fashion but has a long and honourable history and was described and advocated in important publications on the

subject by Hirschfeld in 1935 and Fisher in 1940. The basic method has been "discovered" at least eight times and the currently celebrated version of J.-P. Benzécri is actually only one of the most recent ones. The present success and status of this version owes much to the energy and dedication of Benzécri and his disciples and, in the English speaking and wider world, to Michael J. Greenacre's *Theory and Applications of Correspondence Analysis* (London: Academic Press, 1984) but it seems also to be due in part to a number of other factors such as the following: the recent widespread use and availability of personal computers, particularly the IBM PC and its clones; the existence and availability of a number of relatively simple and friendly programs; the recent growth in and acceptance of multi-dimensional scaling; the discovery and increasing enthusiasm for exploratory data analysis as opposed to classical statistics; and the new interest in graphical forms as ways of presenting the results of statistical investigations.

36. For a fuller account of correspondence analysis at least as applied to literary text see Alastair McKinnon, "Mapping the Dimensions of a Literary Corpus," *Literary and Linguistic Computing* (1989) 73-84. Of course the reader wishing to gain a real understanding of this approach should read Greenacre's book cited in note 35 above.

37. Alastair McKinnon, "Kierkegaard's Attack upon Christendom: Its Lexical History," *Toronto Journal of Theology* (forthcoming).

38. The program also reports an index of concordance which expresses the number of actual common words as a proportion of the number of possible common words and is useful for comparing lists of different lengths. When both lists are the same length the two final scores are for practical purposes identical.

39. Alastair McKinnon, *Dating Kierkegaard's Battles with Fate*, (København, Det Kongelige Danske Videnskabernes Selskab, 1986).

40. All my early results in this study were seriously skewed and showed BI as a far

outlier. Many tests showed that its noun profile was radically different from that of any other work and I finally decided to delete it on the ground that it was his thesis and thus not part of the authorship. I take some comfort in the fact that Kierkegaard himself never treated it as such.

41. Alastair McKinnon, "Mapping the Dimensions of a Literary Corpus," *Literary and Linguistic Computing*, 4, 2, (1989):73-84.

42. The importance of having constant and easy access to the original can hardly be exaggerated. A simple example: many English speaking commentators on Kierkegaard write constantly of his "leap of faith" and the current translation of AE still contains that expression. In fact, a recent search of all the published works has shown that they do not contain a single occurrence of *Troens Spring* its only possible Danish counterpart.