‘I copied all the letters in a big, round hand’: indexing W. S. Gilbert

Geoffrey Dixon

Describes the conception, gestation and birth of the author’s recently published concordance to the Gilbert & Sullivan operas.* The background and history of the project, and the author’s qualifications for the task, are discussed with consideration of the problems facing the concordance-maker—the basic decisions that have to be made and the technical points involved. Base text, arrangement and use, and exclusions are all covered, together with an explanation as to why the concordance was not computer-generated. The mechanics of manual compilation are outlined, as is the preparation of camera-ready copy using a microcomputer-based word-processing program.

Combine a career in librarianship with a life-long interest in music and the theatre (with a particular penchant for the works of W. S. Gilbert and Arthur Sullivan) and what do you get? Resist the strong temptation to reply in Gilbertian manner and what you get in this particular instance is a concordance. A what? A concordance—an alphabetical arrangement of the principal words contained in a book . . . with citations of the passages in which they occur’ (Shorter Oxford English dictionary). This of course makes it an index, and as such of potential interest to readers of this journal. The best-known concordances are those to Shakespeare and the Bible, but many authors have been accorded the honour. For those readers who might wish to refresh their memories as to the appearance of such a work, a glance at figure 1 might be helpful.

A moment’s reflection will prompt the thought that anyone who attempts this considerable labour must be crazy! Why bother? And why me, anyway? The answer to the first question might run something like this: because the Gilbert & Sullivan operas—now 100 years old—have proved to be a truly unique theatrical genre, possessing musical and dramatic characteristics of a quality such as to endear them to a broad spectrum of the public which regards them with love and affection; and also because the operas are the increasing concern of the world of scholarship, as the existence of the W. S. Gilbert Society, the Sir Arthur Sullivan Society, and a growing collection of doctoral theses bear witness. But why me? On account of a lifelong interest as spectator (my first professional Mikado at the age of four), performer, producer and general aficionado. Having played principal parts in, and produced, all the operas, and possessing a not inconsiderable collection of G. & S. material, I have gradually acquired over the years a familiarity with the text of these works which proved invaluable in the concordancing process. So there you have it! A long intimacy and affection combined with a librarian’s compulsion to organize, systematize and make available.

Manual v. computer

Pretty well all concordances nowadays are computer-generated—that is to say, the complete text to be concordanced is typed into the computer using a special program which then manipulates it to produce a context for each word indexed. (These computer-generated contexts are the Achilles’ heel of this method, as will be seen in a moment). Having once typed in the text, therefore, all the drudgery of producing each context, of alphabetizing, and of retyping is eliminated. In this particular case the concordance was first conceived about thirty years ago, and preliminary work was started at that time. Because of this long gestation period the original manual methods were continued. So much of the hard labour had already been done that when the work was resumed after a long lay-off there would have been little point in starting all over again: the point of no-return had been passed. In any case there are certain disadvantages to computer-generated entries. It is obviously impossible to rely on every context formulated by computer software, and a great deal of editing is often necessary. The mere such editing is required, the more the advantages of the computer are nullified. It is no denigration of that monumental work, The Harvard concordance to Shakespeare (Belknap Press, 1973), to point out that such quotations as, ‘now is the winter of our discontent made’ and ‘man he doth bestride the narrow world like a’, do jar a little as a consequence of the unfortunate truncation caused by the demands of the computer program. Manual methods meant freedom from a pre-determined line length. I could give as much—or as little—of the context of any particular word

KISS(n)
with a kiss perhaps on her lantern chaps RG 416
a kiss for that pretty maid YG 450
aye a kiss all round YG 450
a flower and a kiss . . . will bring him round again GD 676
a simple flower a tender kiss will bring him round again GD 676
that flower and that kiss . . . I ne'er shall give again GD 676
that simple flower that tender kiss I ne'er shall give again GD 676

KISS(v)
how dare you kiss me before we are quite married TH 251
bids me humbly kiss your hand SO 31
oh kiss me kiss me kiss me MK 328
you're not going to kiss me before all these people PI 292
mind I will not kiss them MK 329
no no you shan't kiss them MK 329
to mark my admiration I would kiss you fondly thus MK 333
he would kiss me fondly thus MK 333
and on no consideration will I kiss you fondly thus MK 333
now kiss her MK 349
that is where I would kiss her YG 454
oh kiss me kiss me my dear kiss me again and again YG 489
who is about to kiss your hand GN 555
may I kiss him GD 647

KISSED
for he nodded his head and kissed his hand MK 358
they kissed each other's cheek RG 386
oh happy the lily when kissed by the bee RG 408, 432
and fondled and coddled and kissed YG 488 @

KISSES(n)
there'll be bread and cheese and kisses RG 432

Figure 1. An extract from The Gilbert and Sullivan concordance.

as would (a) make sense, (b) preserve Gilbert's rhythms, and (c) render unnecessary, in the majority of cases, further reference to the complete text.

Base text
The concordance-maker must first of all decide upon a base text which he considers sufficiently accurate and complete. The standard edition of the Savoy operas that I used is that published by Macmillan, i.e. Gilbert, W. S. The Savoy operas, 1926 (with frequent reprints). This does not of course correspond in every detail with the text used in stage productions by the D'Oyly Carte Opera Company up to 1982, when the Company was disbanded, since some interpolations and accretions are to be found only in the D'Oyly Carte prompt books, whilst as far as the sung libretto is concerned there are some instances where the vocal score alone contains the quoted words. Conversely, the inclusion in the Macmillan edition of material which has long been traditionally deleted in performance has the advantage of preserving these passages for concordance purposes. In comparative terms the discrepancies are few, and the attempt has been made to include all words customarily used in professional (i.e. D'Oyly Carte) productions in recent decades. Variations from the Macmillan edition are enclosed in square brackets and are the subject of an appended footnote explaining their provenance (see figure 2).

A definitive edition of the Savoy operas, in the sense of a text authorized in total by Gilbert himself, is not really possible. Whilst significant changes were made by Gilbert very early indeed in the runs of many of the operas, subsequent alterations did not all receive his blessing, so that while it would not be true to say that every single word indexed is Gilbert's, the number of such 'rogue' words is very small, and it was felt important to index the text as it is actually known by performers and enthusiasts.

Arrangement and use
For each occurrence of every word indexed a quoted
context is given, usually of a reasonable length, which will allow it to stand alone and retain its integrity as a semantic entity. The reference following each entry consists of two elements: a group of capital letters followed by an arabic numeral. The two letters denote the particular opera from which the extract comes, and of course a key is provided in the prefatory matter. The arabic numerals are the page numbers of the base text. It would have been preferable to use line numbers, but at the time of compilation no line-numbered text was available.

Under each heading the extracts are arranged chronologically: i.e., first of all in the order in which the operas were written, and subsequently, within each opera, in the order of the text. Words which occur more than once on any particular page are again cited chronologically.

Alphabetical arrangement presents few problems in a sequence of single-word headings, and it only became necessary to decide on ‘word-by-word’ in order to deal with hyphenated words, so that love-sick precedes loved.

A major problem for concordances is how to deal with homographs. Does one distinguish between the meanings of, for instance, aye, or between a noun and its identically spelt verb, or does one lump all the contexts together? Similarly with the problem of lemmatization—does one treat every single word as a separate entity or does one, say, list the forms of a verb with its infinitive? And if so, where does one stop? (Ponder for a moment the question of where—and whether—you would look up the word fallen in a dictionary. Answer: you would look up the infinitive, fall.) The purposes of a concordance seem to me best served by ensuring firstly that no word is subsumed under the heading for another, and secondly that the different meanings of homographs are kept apart. The decision was taken, therefore, to distinguish homographs by the addition of an explanatory abbreviation in parenthesis to the heading. It must be recognized, of course, that the renunciation of lemmatization necessarily results in the dispersal throughout the alphabetical sequence of cognate forms, so that, for instance, choose is separated from chose by such words as chop, choral, chords, chorus, etc. Figure 3 illustrates the principle of homograph separation.

A similar problem is presented by contractions and possessives. Normal indexing and filing practice is to ignore the apostrophe in such words as love’s and woo’d, but the fact that this job was being done manually allowed me to separate these pseudo-homographs so that possessives stand apart from simple plurals, and contractions do not become confused with words of different meaning. The rules as eventually formulated were:
1. the possessive singular form files before the plural but after a contraction;
2. the possessive plural form files after the plural;
3. a contraction files before a word formed of the same letters but lacking the apostrophe;
4. in all other cases (e.g. ne’er) the apostrophe is ignored, thus:

Excluded words

Even in a computer-generated concordance it is not usual to include words such as a, and, not, etc., and a decision had to be taken as to which words to exclude. In an effort to make the ‘stop list’ as short as possible, the following were the main categories of excluded words:

- articles;
- the commoner prepositions (but quite a few were in fact indexed);
- pronouns (but possessive and reflexive pronouns were indexed);
BOUND\(\text{S}(\nu)\)
over the hedges and brooks she bounds

BOUQUET
her bouquet is simply frightful

BOW\((n)\) (FRONT)
I'm all bow though I'm sure I try to be stern

BOW\((n)\) (OBSEQUY)
a little bow—a mere nothing
this deathly bow was a touching sight to see

BOW\((n)\) (WEAPON)
as the modest moon with crescent bow
two strings go to every bow

BOW\((v)\)
he never should bow down
to thy behest offended Queen I bow
bow bow ye lower middle classes
bow bow ye tradesmen bow ye masses
bow thy head to Destiny
to your command on every hand we dutifully bow
list to their aims and bow your head in wonder
at my exalted name Posterity would bow in gratitude
bow bow to his daughter-in-law elect
let us bow to Fate's decree
and noble lords will scrape and bow
bow impressively ere you glide
with loyalty blind I [bow me] to thee

Figure 3. Separation of homographs.

possessive adjectives;
conjunctions;
negatives no and not;
auxiliary verbs be, can, have, etc.;
demonstratives;
and a few other odds and ends.

Everything else was indexed—hence the 63,000 or so entries.

Alliteration

One characteristic of Gilbert's style is his widespread use of alliteration. Well known instances include the 'short, sharp shock' lyric from *The Mikado*, and Casilda's 'Papa I've no patience with the presumption of persons in his plebeian position' from *The Gondoliers*, but Gilbert's use of this device is extensive—really quite remarkable—and often goes unnoticed, concealed by the felicity of his construction and phrasing. It was considered useful therefore to signal all instances of alliteration, and this was done by the use of the symbol @, e.g.:
dark danger hangs upon the deed YG 443 @

Compilation

Bearing in mind, then, that at the inception of this project we were in the era BC (before computers), the method adopted was roughly as follows. Starting at the beginning of *Trial by jury* the text was transcribed—in manuscript—phrase by phrase on to 5" × 3" slips. (Note for Savoyards: the first collaboration of G. and S.—*Thespis*—is not included in the Macmillan edition and was printed as an appendix to the concordance and indexed to its line numbers.) The words in each phrase which were to be used as headings were ticked on the slip, and from the original slip the required number of copies was made, each with its own appropriate heading. Thus a context which incorporated four indexable words would have three additional copies written out with the correct heading added. This meant that the important question of exactly what context to quote was mine to decide, and the copying of the required number of additional slips could be left to slaves (both paid and unpaid!).

The completed slips were left in the order of the text so that a check could be made for correct page numbers.
When all the text had been transferred to around 63,000 slips—and we are skipping lightly here over a time span measured in years—the next task was to check the base text (Macmillan edition) against other editions including the vocal scores in order to record any discrepancies which were to be the subject of footnotes in the concordance. Incidentally, this was done by reading the text aloud to two or three helpers, each of whom had a copy of a different edition. All that then remained was to convert the contents of 24 catalogue drawers from chronological to alphabetical order, and to arrange all the slips under each heading into chronological order. This last was not too difficult, since the Macmillan edition prints the opera texts in the order in which they were written, and therefore a sort by page number was all that was required. (Arrangement within each page number meant either relying on my own knowledge of the text or on checking the actual page.) At this point the various cross-references which had accumulated were interfiled. In comparative terms these were few, being mainly 'see also' references from everyday terms to their poetic, archaic or dialectal equivalents (e.g. before see also ere; knew see also knewed) but including 'see' references from forms not used (e.g. loath see loth).

**Publication**

Since Macmillan are the publishers of the base text and also specialize in large works of reference it seemed obvious that my first approach should be to them. However, after an interesting and friendly correspondence with Lord Macmillan himself (as he was then), they felt that they had to refuse. Eventual success with Garland Press brought a requirement for camera-ready copy; which of course meant an immaculate, error-free typescript. The plan of campaign called for me to type the work myself, since by so doing I could effect any necessary editing as I went along, armed as I was with a fairly intimate knowledge of the text. This in turn pointed to a word-processor and, because it happened to be convenient, the EDWORD chip in the BBC micro was used. In spite of the fact that this generated around 400 files on 40 discs it did a reasonable job, and the resultant print-out from a daisy-wheel printer was published after photographic reduction from A4.

The Gilbert and Sullivan operas are by now a national institution as well as constituting an increasingly significant field of academic study. (What would W. S. Gilbert have made of that?) They are of sufficient literary merit to deserve such attention and effort. I kept no record of the number of hours devoted to this project since its inception, but I do know that, two volumes and 1,780 pages later, I would happily do it all over again.

---

**Geoffrey Dixon is Principal Librarian of Craigie College of Education, Ayr; editor of and contributor to the current course on indexing offered by the Rapid Results College; volume indexer for The Indexer.**

### Primary/Secondary Information Services Co-operation

To strengthen the relationship between primary literature producers and secondary information services, they say, the National Federation of Abstracting and Information Services (NFAIS) has produced a for-members-only information pamphlet entitled *Primary and secondary information services: basic information from NFAIS* that explains how these two types of information organizations complement one another.

The principal objective of both primary publishers and secondary information services is the same—to facilitate the transfer of information. Primary publishers provide the complete, original information in a reviewed, edited, readily available, archival form. Secondary information services provide means for easily accessing the original literature. Because of the complementary nature of their relationship, each benefits from co-operation. The pamphlet explains how this can be achieved, ways that primary producers and secondary services are already working together, CODEN and ISSN numbers, the history of NFAIS, and how journal publishers benefit when their titles are indexed by secondary services—such inclusion often being a criterion for journal acquisition.

---

**The indexer as megalomaniac?**

A whole new field of objection to wifely indexing—and editing—becomes apparent in an interview by Stephanie Mansfield with Patricia Cotten, wife of the film actor Joseph Cotten, in *The Age*, 12 Sept. 1987, discussing his biography, *Vanity will get you somewhere* (reported in the AusSI Newsletter, August 1987). It goes:

‘He wanted to be taken as an author,’ says Patricia, lashes fluttering. It was she who edited the book. In fact, she says, after her husband’s stroke he was unable to work, so she cut 500 pages from the manuscript. What’s left is an index somewhat short on references to Ingrid Bergman, Alfred Hitchcock and Marilyn Monroe, just a few of the luminaries with whom Cotten worked closely. Instead, we find information on Patricia herself: ‘Medina, Patricia: beauty described ... as card player ... courting of ... see also Cotten, Patricia Medina ... appreciation for ... care during illnesses ... as coach ... early medical aspirations ... honeymoon with ... meeting with Hepburn ... orthopaedic problems ... photographic memory ... remodelling recommendations ... after Selznick’s death ... in Venice ... willingness to travel.’

Is Cotten that coy about his colleagues, or did his wife of 26 years edit them out? ‘I don’t like dirt really,’ he explains.

---

**The Indexer** Vol. 16 No. 2 October 1988