Subheadings in award-winning book indexes: a quantitative evaluation

Cecelia Wittmann

The subheadings in four indexes, winners of the Wheatley Medal, were analyzed in an attempt to quantify those characteristics of subheadings that make good indexes. This analysis indicated that there are significant quantitative differences between award-winning and ordinary indexes, though the characteristics of award-winning indexes vary with the nature of the text. Award-winning indexes to historical narratives were found to have subheadings about 5 words in length; most of these subheadings began with a significant word, either a noun or a verb; most were not related syntactically to their main headings; and only about 20% matched the words of the text. These characteristics were not shared by ordinary indexes.

Books and articles on indexing agree that subheadings are essential in an index, but give little exact information about what makes subheadings good. The most complete and useful information extant still seems vague and inadequate. The ANSI standard for indexes (Z39.4-1984, Basic Criteria for Indexes) states that subheadings should begin or end with prepositions or conjunctions only if necessary for clarity: for example, ‘Bibliographies: scope’, not ‘Bibliographies: scope of’. Indexing concepts and methods, by Harold Borko and Charles Bernier, states that subheadings should start with ‘unusual and eye-catching words’, which could be adjectives only if they are ‘exceptional or unique’. The Chicago manual of style, without giving further details, states that subheadings must be related logically to headings, and sometimes can be related grammatically, so that the heading and subheading combined form a meaningful phrase, such as “Statistical material: units of measure in”.

A cursory comparison of several ordinary indexes with indexes that had won the Wheatley Medal suggested that subheadings in award-winning indexes are vivid and concise, conveying in a few words the essence of the material indexed; in contrast, the subheadings in other indexes were often cryptic, rambling, or vague. Could the clear superiority of subheadings in award-winning indexes be explained solely in terms of their having the characteristics advised by sources such as those listed above? Or did it result from additional qualities?

I undertook a detailed analysis to determine whether or not this subjective impression of qualitative differences could be formulated on the basis of such objective and quantifiable features as subheading length, initial word of subheadings, syntactic relationships between subheadings and main headings, and semantic relationships between subheadings and the text indexed. I felt this was important, since such information about the characteristics distinguishing high-quality indexes from mediocre ones could form the basis of more definite criteria for evaluating indexes. In addition, it could be useful in teaching indexing courses as well as in helping professional indexers improve the quality of their indexes.

In this analysis, four indexes that had been awarded the Wheatley Medal for indexing were paired with four apparently good-quality indexes in books on similar topics. All eight indexes had run-on subheadings, since a preliminary comparison of award-winning indexes suggested that run-on subheadings and set-out subheadings have characteristically different length and syntax.

All four award-winning indexes were published within a span of only six years, so any differences in them could not reasonably be attributed to changing fashions in indexing style. In addition, since these works all deal with history and biography, they had similar indexing requirements (that is, names, places, and events), and therefore formed a unified group for analysis.

Indexes in the following four pairs of works were selected for analysis and comparison:


SUBHEADINGS IN AWARD-WINNING BOOK INDEXES


The number of subheadings to be analyzed from each index was determined using a statistical formula to obtain an adequate sample. The specific subheadings to be analyzed were then selected randomly, from 55 to 149 per index, as follows:

Winners of the Wheatley Medal:
'Clemency' Canning: 91
Principal Navigations: 60
Wardens' Accounts: 55

Other indexes:
Winston S. Churchill, vol. 3: 100
Sir Charles Wood: 149
Voyages: 100
Acts of Court: 61

Each subheading chosen, along with its corresponding main heading and page references, was written on a 3" x 5" card, and the following information was then written on each card for analysis:

1. Number of words in the subheadings (excluding dates given in parentheses)
2. Part of speech of the first word in the subheading
3. Semantic relationship of the subheading to its main heading:
   a) Direct Natural reading order of the phrase is main heading followed by subheading:
      Muhammadans in 'native' regiments
   b) Indirect Natural reading order is the subheading first, followed by the main heading:
      Education Bill religious schools threatened by
   c) Topical Main heading and subheading do not together form a phrase:
      Diseases, remedies for palm oil, antidote to poisoned arrows
4. The relationship of the subheading to the text:
   a) Exact match All words of the subheading match, in order, the text phrase indexed.
   b) Near match The words of the subheading and text differ only in singular or plural, or in prepositions.
   c) Synonymous match One or more words in the subheading is a synonym of a word in the text phrase indexed.
   d) Paraphrase The subheading has few or no words of the text indexed, or contains words from the text plus additional words supplied by the indexer.
5. An indication of which words in a subheading occur in the text indexed.

Results and discussion

A preliminary subjective examination of several indexes had suggested that the subheadings of award-winning indexes were generally shorter and more vividly descriptive than those in ordinary indexes. The present analysis was undertaken to determine whether or not this subjective impression of differences could be verified by such objective and quantifiable features as subheading length, initial word of subheadings, syntactic relationships between subheadings and main headings, and semantic relationships between subheadings and text.

The analysis found that award-winning indexes to historical narratives had similar features, and differed significantly from an award-winning index to a documentary text and from non-award-winning indexes. The results, summarized in Table 1, are discussed below.

Table 1. Summary of differences among indexes analysed

<table>
<thead>
<tr>
<th>Subheading feature</th>
<th>Award-winning narratives</th>
<th>Non-award-winning narratives</th>
<th>Historical narratives</th>
<th>Mercers'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average no. of words</td>
<td>4.7–5</td>
<td>2.3</td>
<td>3.4–7.5</td>
<td>2.6</td>
</tr>
<tr>
<td>Order direct</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>indirect</td>
<td>30–34</td>
<td>7</td>
<td>10–42</td>
<td>41</td>
</tr>
<tr>
<td>topical</td>
<td>19–40</td>
<td>9</td>
<td>38–62</td>
<td>41</td>
</tr>
<tr>
<td>30–48</td>
<td>84</td>
<td>12–28</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>First word n., v., prep.</td>
<td>60–68</td>
<td>87</td>
<td>38–82</td>
<td>74</td>
</tr>
<tr>
<td>adj. or adv.</td>
<td>10–27</td>
<td>7</td>
<td>6–14</td>
<td>7</td>
</tr>
<tr>
<td>conj., art., or prep.</td>
<td>9–20</td>
<td>13</td>
<td>12–70</td>
<td>20</td>
</tr>
<tr>
<td>Relation to text exact or near match synonymous</td>
<td>18–23</td>
<td>55</td>
<td>9–16</td>
<td>36</td>
</tr>
<tr>
<td>paraphrase</td>
<td>2–15</td>
<td>11</td>
<td>11–14</td>
<td>25</td>
</tr>
<tr>
<td>65–80</td>
<td>35</td>
<td>71–77</td>
<td>39</td>
<td></td>
</tr>
</tbody>
</table>

Note: Since percentages have been rounded to the nearest whole number, some totals may be more than 100%.
Length of subheads: The award-winning indexes to the three historical narratives have subheading lengths averaging 4.8 words. Although the subheadings range from 1 to 14 words long, the great proportion are in the narrow range from about 3 to 5 words long. In contrast, the average length of subheadings in non-award-winning indexes ranged from 3.4 to 7.5 words, with a significant number ranging from a terse 1 word to a rambling 10 words. The longest was 19 words. This suggests that the optimal length of subject subheadings for historical narratives is about 5 words, which enables the subheading to describe concisely the text indexed, while including enough description to let the reader decide whether the text would be of interest. Subheadings much shorter than 5 words may be too brief to convey adequately the content of the text indexed, while subheadings that are significantly longer may be rambling and not necessarily clearer.

The subheadings in Wardens’ accounts are significantly shorter: about 2.25 words. This may be because the text is principally a collection of legal and other documents, and the index contains a large proportion of one- and two-word subheadings for the various objects mentioned in these documents. This fact suggests that the ‘best’ subheading length may be different for different types of texts.

Order of subheadings. In all the award-winning indexes, topical subheadings predominate; that is, the subheadings are not syntactically related to the main heading under which they occur. For example:

Bengal army
enlistment terms

Approximately 30% of the subheadings in indexes to historical narratives and 7% of the subheadings to Wardens’ accounts are direct-order subheadings—the heading and subheading, read in that order, form a natural phrase, as in:

Norwegians
take British wives

In contrast, the majority of the subheadings in the non-award-winning indexes are indirect-order subheadings, in which the subheading followed by the heading forms a phrase, when read in that order:

Novaya Zemlya
latitude of

Topical subheadings, the least frequently used, constitute 12–28% of the total. These differences suggest that subheadings in award-winning indexes generally do not end with prepositions; that is, they are topical subheadings (such as ‘Lucknow: military problems posed by capture’), not indirect-order subheadings (such as ‘Marseilles: weights of’). Whereas these end with prepositions only when needed for clarity, the subheadings in non-award-winning indexes frequently end with nonessential prepositions.

Initial word of subheadings. All the indexes differed in the proportion of nouns, participles, and other types of words constituting the initial word of subheadings. However, award-winning indexes differed significantly from the other indexes in their proportion of descriptive words (nouns, verbs and participles) used as initial words in subheadings. The award-winning indexes, excluding Wardens’ accounts, had similar proportions of subheadings beginning with nouns, verbs and participles—about 66%. (In the index to the latter, which is principally a collection of documents, 84% of the subheadings start with nouns, verbs or participles.) Further, only about 15% of the subheadings in the award-winning indexes begin with such nondescript words as articles, conjunctions, and prepositions.

The subheadings in the non-award-winning indexes were more variable: from 38% to 82% begin with content-rich nouns, verbs and participles, and from 11% to 70% begin with articles, conjunctions, or prepositions. The difference between that two groups of indexes is especially striking with regard to conjunctions. They are the initial word in very few of the subheadings in the award-winning indexes, but are relatively common in the other indexes. Almost twice as many subheadings in the award-winning indexes, compared with non-award-winning indexes, begin with potentially descriptive, content-rich adjectives and adverbs.

These facts suggest that the subheadings in award-winning indexes are more consistently content-rich than the subheadings in other indexes, if it can be assumed that the proportion of descriptive to nondescriptive words at the start of a subheading is an indicator of richness of content.

Direct use of author’s words. Differences in the extent to which the words in the subheadings were drawn directly from the text correlated principally with the type of text indexed, historical narrative or documentary. Award-winning and non-award-winning indexes were generally similar. For indexes to historical narratives, about 20% of the subheadings in award-winning indexes and an average of 14% of the subheadings in the other indexes, were exact or near matches of the text indexed. In contrast, subheadings in indexes to documentary material were exact or near matches about 2½-times as often. They were 55% of the subheadings in the award-winning Wardens’ accounts, and 36% of subheadings in Mercers’ Company. Paraphrases accounted for about 70% and 74% of the subheadings in award-winning and other indexes, respectively, to historical texts, but were about 37% in indexes to documentary material.

Conclusion
This analysis indicates that the characteristics of good subheadings vary according to the nature of the text indexed.

Subheadings in indexes to historical narratives average...
about 5 words in length, whereas those to texts that consist principally of documents are about half as long. This difference is probably because the former are phrases summarizing an event described in the text, whereas many of the latter are terms for objects and occupations named in the text.

About 40% of the subheadings in indexes to historical narratives were topical; that is, they are not syntactically related to their main headings. In contrast, 84% of the subheadings in the index to a documentary text are topical. This difference reflects the nature of the indexing phrases used: in historical narratives, each is often a complete phrase, part of which forms the main heading and part the subheading; in documentary texts, on the other hand, there are few phrases, since many subheadings are narrow terms under broad main headings.

An overwhelming majority of the subheadings in award-winning indexes began with such significant words as nouns, verbs or participles. The proportion is highest in indexes to documentary material, since many of the subheadings are nouns describing occupations or things.

The relationship between the words of the subheadings and the words of the text varies according to whether the index is principally to events and ideas or to persons and things. In the latter case, such as an index to a documentary text, where the actual words in the documents are often significant, the majority of the terms in the subheadings exactly or nearly match the words used in the text. In contrast, only 20% of the subheadings in indexes to historical narratives closely match the text, probably because the task of the indexer here is principally to summarize the various events and ideas described in the text, not to provide access to the author's own words. The proportion of subheadings that contain synonyms or paraphrases of an author's words may differ with the style of the author: the indexer may be able to use the exact words of authors whose words are vivid and highly descriptive, but may choose to paraphrase authors whose vocabulary is less picturesque.

To what extent do the subheadings in award-winning indexes follow ANSI and other recommendations? The ANSI standard states that prepositions and conjunctions should be at the beginning or end of subheadings only when necessary. Borko and Bernier state that the initial word of a subheading should be 'unusual and eye-catching'; that is, not a preposition or conjunction. Good indexes have the characteristics recommended by the ANSI standard, but not those recommended by Borko and Bernier: up to 20% of the subheadings start with such nondescriptive words as prepositions and conjunctions. It is possible that the recommendations of Borko and Bernier more closely describe good subheadings in the set-out (line-by-line) format, in which each subheading starts a new line. This format enables rapid scanning of the initial word of the subheadings.

In contrast to the subheadings in high-quality indexes, many subheadings in the other indexes begin or end with prepositions or conjunctions. This may contribute to the longer average length of these subheadings, in comparison with subheadings in award-winning indexes.

Because of the small number of indexes analyzed, the results of this study are only preliminary. Nevertheless, they suggest that it is possible to quantify some of the characteristics of subheadings in indexes, and to find statistically significant similarities among these characteristics in high-quality indexes to similar types of texts. This finding suggests that indexing standards and guides—whether dealing with depth of indexing, structure of headings and subheadings, or other features—should be developed around particular types and subjects of texts, rather than attempting to define single standards to characterize all indexes.

The author wishes to thank Dr Elaine Svenonius for assisting in design of this study and for making helpful comments on earlier drafts of this article.

References

Cecelia Wittmann is on the adjunct faculty of the University of California at Los Angeles, where she teaches indexing; and of San Jose State University, where she teaches cataloging and classification. She is Recording Secretary of the American Society of Indexers.

Sickly circle

'The publications are not as well known as they deserve to be because they have not been adequately indexed, while they have not been indexed because they are not well known.'