The definite article: acknowledging 'The' in index entries

Glenda Browne

This article examines rules and practice relating to the filing of 'The' at the beginning of index entries, recommending that the definite article should be accorded more respect, and be considered in filing when the nature of the index entry warrants it.

> Happy is the lot of an indexer of Latin, the Slavic languages, Chinese, Japanese, and some other tongues which do not have articles, whether definite or indefinite, initial or otherwise. (Wellisch, 1991: 188)

'The' is an unusual word. It is crucial, and common, yet it is regularly omitted or ignored. What should indexers do about it? How, for example, should we index 'The Who' or 'The Hague' or *The Lion, the Witch and the Wardrobe*?

Based on principles of user expectations, simplicity and consistency, I suggest that we should pay more attention to 'The', acknowledging its role as a significant word. If 'The' exists in a name or title, it should exist in the index entry for that name or title. And if it exists in the index entry, it should be taken into account when sorting the entries.

Rules about 'The'

When creating an index entry, indexers consider two things: what is the accepted rule for this situation, and where might users look for information on this topic. In English works, the basic rule is to ignore 'The' when indexing titles of publications and corporate names, but to take it into account when indexing first lines of poems and names of places. When it is ignored it may be omitted entirely, inverted, or retained but ignored in sorting.

Titles starting with 'The'

The Australian/New Zealand Standard (AS/NZS 999, 1999), the text of which is identical to the international standard (ISO 999), says:

> In an English index, articles in titles are conventionally transposed to the end of the heading so that filing order is explicit.

The *Anglo-American Cataloguing Rules* (AACR, 1988) advise omitting initial articles from uniform titles (25.2C1) and from titles appearing in references (26.1A). Nancy Mulvany (1994: 122–3), in *Indexing books* states:

> If the first term [in a main heading] is an article, such as *A* or *The*, the term is not alphabetized. For example, the book title *A Tale of Two Cities* would appear in the *T's*, sorted on 'Tale'. [But] if an article, preposition, or conjunction appears in any other position besides as the first term in an entry, it is always alphabetized.

In *Indexing from A to Z,* Hans Wellisch (1991: 188) says:

Because of the very high frequency of initial articles in titles of works, it has for a long time been the rule to transpose such articles to the end of the title, preceded by a comma, so that the title can be filed by the (capitalized) word following the article, e.g. *Tempest, The.*

In *Indexes and indexing*, Collison (1972: 198) advises us:

> Ignore definite and indefinite articles in arrangement. Omit them where possible; where it is necessary to retain them, invert: *Times, The*; *Temps, Le*.

The index to his book has an entry 'Bible, The', showing that he felt that 'The' could not be omitted there, even though it was not capitalized in the text.

Place names starting with 'The'

For place names beginning with articles, such as The Hague, Mulvany (1994: 175) says that the article 'The' is part of the formal name of the place, and appears on maps. Therefore:

> The names are entered in the index in uninverted form, sorted on *The*. It might be very helpful to readers to also supply cross-references, as in 'Hague *see* The Hague'.

Similarly, the Australian/New Zealand standard (AS/NZS 999, 1999) says:

> An article or preposition should be retained in a geographic name of which it forms an integral part.

First lines starting with 'The'

According to Wellisch (1991: 189):

> An exception to the rule of transposition of initial articles has always been made for indexes of first lines of poems and anthologies. All such initial articles are printed and filed as any other words, followed by the rest of the first line (which may also be the title of the poem).

Corporate names starting with 'The'[1]

Wellisch (1991: 190–1) says that:

> Initial articles in corporate names and in proper names of ships, cars, airplanes, buildings, etc. are transposed, e.g., . . . West Bend Co., The.

This rule also includes foreign names, such as 'Trattoria, La' (Wellisch, 1991: 191).

The Anglo-American Cataloguing Rules (1988) say:

> 24.5A1. Omit an initial article unless the heading is to file under the article (e.g., a corporate name that begins with an article that is the first part of the name of a person or place).

The examples it gives are to use *Library Association*, not *The Library Association*, but to use *Los Angeles Symphony* (with the article *Los*) because it starts with the name of a place. The Australian/New Zealand Standard also agrees that 'an initial article is omitted, unless specifically required'.

What really happens with 'The'

Titles starting with 'The'

While 'The' is often inverted or omitted from titles, this is not always the case. The index to *Sitting on penguins* (Murray-Smith, 1991), for example, sorts titles starting with 'The' under 'The'. Index entries include '*The Home of the Blizzard* (by Douglas Mawson)' and '*The Worst Journey in the World* (by Apsley Cherry-Garrard)'. These titles are not doubled under 'Home' and 'Worst'. The indexer is not acknowledged, but the author, Stephen Murray-Smith, is a literary figure in Australia.[2]

In the list of journals on the Stockton Press website,[3] *The Hematology Journal* files after *Spinal Cord*. Presumably simple computer filing order has been followed.

In the General Index to *The Australian Writer's Marketplace 1999* (Whitton, 1998), the newspaper *The Australian* is entered as *Australian, The*, and has been automatically sorted between *Australian Tennis Magazine* and *Australian Theatre for Young People*. Once again computer sorting appears to have taken over. In this case, the inversion of 'The' makes the computer sort order totally inappropriate. Since 'The' really belongs at the beginning of the title, it is wrong to sort on it as the second word in the title.

Interestingly, in *Sitting on penguins, The Australian* is entered as *Australian* (newspaper) without any 'The' at all. When it is referred to in the text 'The' is not capitalized, indicating that it has not been considered an integral part of the title. On the other hand, the company that publishes the newspaper is listed in the Sydney telephone directory as *The Australian*, sorted under 'The'.

Place names starting with 'The'

In the *UBD Sydney Professional Drivers Street Directory* (1998), streets such as *The Corso* and *The Northern Road* are sorted under 'The'. This follows the rule of indexing places without inversion.

In the Sydney (NSW) telephone directory (2000/2001), corporate bodies with names starting with the full name of a place starting with 'The' are sorted under 'The' (e.g. *The Rocks Teppanyaki* and *The Oaks Police Station)* or under the second word of the place name (e.g. *Rocks Chamber of Commerce Inc The)*. No comma is used between 'Inc' and 'The' to show that it is an inversion.

Corporate names starting with 'The'

The introduction to the 2000/2001 Sydney (NSW) telephone directory states: 'Business names which begin with "The" are generally sorted under the next word. **Check both places**' [their emphasis]. The newspaper *The Age* has entries at *The Age* and *Age The* (in the 1997/1998 directory there were fewer phone numbers at *The Age* than at *Age The)*. *Agency Register The* and *Agency Personnel The* are sorted under 'A', while *The Agency Australia* is sorted under 'The'. *The Sausage Specialist* is filed under 'The' and 'Sausage', while *The Meat Emporium* is only filed under *Meat Emporium The*. The user's only option is to look in both places.

What do users think?

Without doing extensive user testing, it is difficult to say where users would look; however, we can assume that many will look where the rules suggest they should, and others will look in the places where they find information in other sorts of documents. When talking with indexing and non-indexing colleagues I have found no consistent approach to 'The'.

The nature of 'The'

There are two reasons why we have such trouble agreeing what to do about 'The'. First, it is so common that in certain lists it has been considered more useful to promote the second word. This is particularly true of library catalogues where many book titles start with 'The'. This reason is insignificant in most indexes.

The other reason is that 'The' often doesn't matter. There are many titles that include 'The', but then treat it as if it doesn't exist. The masthead of *The Australian*, for example, has a tiny *The* above a large *Australian*. Their layout tells us that *The* is insignificant, but they won't follow this through and omit it entirely.

Corporate names such as *The University of Queensland* are used at times with, and at times without, an initial 'The'. This makes it very difficult for users to know whether 'The' is an integral part of the name, and is a major reason why it is ignored in the filing of corporate names. John Kerr (2000: 5), an editor, has written about 'the Pretentious The', complaining about the unnecessary and ugly use of *The* in *The Courier-Mail* and *The University of Queensland*.

We have *The Mount Sinai Journal of Medicine, The American Journal of Psychiatry* and *The Journal of Molecular Diagnostics*, but *British Journal of Pharmacology* and *Journal of Microscopy*. There is no significance in the presence or absence of 'The'. Some journals consider 'The' to be an integral part of the name, as seen in the email address of the journal *The Lancet*, which includes 'the' (www.thelancet.com).

On the other hand, in many corporate names 'The' has been deliberately chosen as the first word of the name, and is used consistently. The musical group *The Beatles* is referred to as such, and never as *Beatles*. In these cases, the group considers the initial article significant, and it will be the access point consulted by many users. An extreme example is the group *The The*, which would look absurd with the initial 'The' omitted or inverted. (Although, admittedly, it would be easily retrieved no matter which sorting method was used.)

Another example of a significant 'The' was given by Priscilla S. Taylor (2000), who wrote:

Several sharp-eyed readers of the newsletter I edit alerted me to the fact that the title of Darwin's pathbreaking work is *The Origin of Species* (not *Origin of the Species*, as it had appeared in the newsletter), and that misplacing the in this instance is meaningful and has deeply disturbing implications.

What should indexers do about entries starting with 'The'?

Indexers apply generally accepted rules of indexing and also use individual judgement, always thinking: 'Where would a user look for this item?'. Since some users do not know the rules, I suggest we should put entries starting with 'The' wherever those users might look; that is, both sorted on 'The' and as a double entry under the second word. Where the length or complexity of the index is a factor, then a cross-reference would replace the double entry, and the indexer would have to decide which form to prefer.

Reasons for sorting on 'The'

Taking 'The' into account when sorting would be useful for the following reasons:

- We would have consistency between all the rules, rather than inverting in titles and corporate bodies, and leaving 'The' at the beginning of places and first lines of poems.
- We would have consistency within one title. Currently initial articles are ignored, but articles in the middle of a title are taken into account when filing. This is not intuitive. In the book *The lion, the witch, and the wardrobe*, for example, each 'The' is as important as the others. If we sort on the initial 'The' (as well as the following ones in their turn), then we are according each of the articles equal importance.
- We would follow the modern trend of sorting things as they are naturally, rather than using intricate filing rules to re-sort entries. No longer do we sort '1000' under 'o' for 'one thousand', nor do we sort 'St Anne' as 'Saint Anne'.[4]
- We acknowledge the words that an author has used; if 'The' is in their title, then it should be accorded significance in our index.

The same applies, and even more so, to articles in other languages. An Australian is very likely to look for the show *Les Miserables* under 'L'. To us, *Les* is an integral part of the name, and not to be discarded. (In fact, we often abbreviate it to *Les Mis*, so *Les* is more important to us than '*erables*'.) Even users who know the rules about articles don't necessarily know which foreign words are articles.

Reasons for not sorting on 'The'

The reasons for not sorting on 'The' are:

- Users who know the current rules will not look for a title or corporate name under 'The'.
- Some indexes will contain long lists of titles starting with 'The', making specific entries harder to find.
- People often do not know that the name of the thing they are looking up contains 'The'.

- Similar entries based on place names but without 'The' will not sort together (for example, *The Corso*, and *Corso Restaurant*). In fact, these do not sort together under the current rules either.

Double entry with and without 'The'

There are good reasons for sorting on 'The', and good reasons for ignoring it. The win–win situation is therefore to have double entry of titles, place names and corporate names under 'The' *and* under the second word in the entry. If this creates unmanageably long lists of entries starting with 'The', then a reference could be added. For example:

> The . . ., to search for titles and names starting with 'The', *see the second word in the entry.*

This quickly lets users know what rule you are using. Alternatively, information about the way you have dealt with 'The' can be included in the introduction to the index.

While references like these might look wordy for our tidy indexes, one colleague has said to me that she is thinking it is better to make our indexes less pristine, and more helpful. A side benefit of this is that the more guidance we put into indexes, the more likely users are to realize that people create them, and to recognize the value that human indexers can add. A brief index with no guidance may be concise, but it is not necessarily the best help to the user.

Conclusion

If we strictly follow indexing rules about inversion of 'The', we run the risk that some of our users will not find the information they are looking for. If we look at a range of information sources we find that there is no consistent approach. Information creators deal with it differently, and users have to cope with different approaches. As indexers, we can ensure that users find what they are looking for in our indexes by making sure that we double entries at 'The' and at the second word in the entry. Similar arguments apply to 'A' and 'An' but these are beyond the scope of this article.

Notes

1. Corporate names refer to organizations or groups of people acting as an entity, e.g. companies.
2. The blurb to *Sitting on penguins* notes that Murray-Smith was a professional historian, editor of the literary magazine *Overland* and of *The Dictionary of Australian Quotations*, and author of *Right Words*, a guide to English language usage in Australia.
3. http://www.stockton-press.co.uk/journals/ (Accessed 16 June 2000).
4. Mulvany (1994: 133) says that many scholarly publishers require that numerals be alphabetized as though spelled out but that 'a more sensible approach is to arrange numerals in ascending numeric order', while Wellisch (1991: 268) says that 'the solution . . . is the arrangement of all numbers written in numerals in ascending order after spaces . . . and before any heading beginning with a letter.' Wellisch (1991: 3–4) is also an

advocate of filing abbreviations for Saint (St, Ste) as they are given, rather than spelt out as Saint, Sainte, Sankt, Santa, Santo and so on and Mulvany (1994: 158) says that 'sorting abbreviated "St." names as if they are spelled out is yet another quirk that has become common practice. Anytime an alphabetic list is not sorted alphabetically we must take pause.'

References

AACR (1988) *Anglo-American Cataloguing Rules,* 2nd edn, Michael Gorman and Paul W. Winkler, eds. Chicago, IL: American Library Association.

AS/NZS 999 (1999) *Information and documentation – Guidelines for the content, organization and presentation of indexes* (ISO 999: 1996). Homebush, NSW: Standards Australia.

Collison, Robert L. (1972) *Indexes and indexing*. London: Ernest Benn.

Kerr, John (2000) Pretentious titles. *Offpress: Newsletter of the Society of Editors (Queensland) Inc.* June 2000.

Mulvany, Nancy (1994) *Indexing books*. Chicago, IL: The University of Chicago Press.

Murray-Smith, Stephen (1991). *Sitting on penguins: Australia and the Antarctic*. Port Melbourne, Victoria: Mandarin Australia.

Taylor, Priscilla S. (2000) How careful should editors be? *The Society of Editors (Victoria) Newsletter* **29**(7) (previously published in *The Eye*).

UBD (1998) *UBD Sydney Professional Drivers Street Directory,* 34th edn. Macquarie Centre, NSW: Universal Press.

Wellisch, Hans H. (1991) *Indexing from A to Z*. New York: The HW Wilson Company.

Whitton, Rhonda, ed. (1998) *The Australian Writer's Marketplace 1999: The complete guide to being published in Australia*. Melbourne, Victoria: Bookman Press.

Glenda Browne trained as a biotechnologist and librarian. She has indexed books, journals, websites, databases and online help for 11 years. With Jonathan Jermey she has written the soon-to-be-published book Website indexing. *She was a founding committee member of the NSW Branch of the Australian Society of Indexers and was the AusSI newsletter editor from 1998 to 2000. Email:* `webindexing@optusnet.com.au`

The Dogge called the Setter

Another sort of Dogges be there, seruicaeble for fowling, making no noise either with foote or with tounge, whiles they followe the game. These attend diligently vpon theyr Master and frame their conditions to such beckes, motions, and gestures, as it shall please him to exhibite and make, either going forward, drawing backeward, inclining to the right hand, or yealding toward the left, (In making mencion of fowles my meaning is of the Partridg and the Quaile) when he hath founde the byrde, he keepeth sure and fast silence, he stayeth his steppes and wil proceede no further, and with a close, couert, watching eye, layeth his belly to the grounde and so creepeth forward like a worme. When he approcheth neere to the place where the birde is, he layes him down, and with a marcke of his pawes, betrayeth the place of the byrdes last abode, whereby it is supposed that this kinde of dogge is called Index, Setter, being in deede a name most consonant and agreable to his quality. The place being knowne by the meanes of the dogge, the fowler immediatly openeth and spreedeth his net, intending to take them, which being done the dogge at the accustomed becke or vsuall signe of his Master ryseth vp by and by, and draweth neerer to the fowle that by his presence they might be the authors of their owne insnaring, and be ready intangled in the prepared net, which conning and artificiall indeuour in a dogge (being a creature domesticall or householde seruaunt brought vp at home with offalls of the trencher and fragments of victualls) is not much to be maruailed at, seing that a Hare (being a wilde and skippishe beast) was seene in England to the astonishment of the beholders, in the yeare of our Lorde God, 1564 not onely dauncing in measure, but playing with his former feete vppon a tabbaret, and obseruing iust number of strokes (as a practcioner in that arte) besides that nipping & pinching a dogge with his teeth and clawes, & cruelly thumping him with y'force of his feete. This is no trumpery tale, nor trifling toye (as I imagine) and therefore not unworthy to be reported, for I recken it a requitall of my trauaile, not to drowne in the seas of silence and speciall thing, wherein the prouidence and effectuall working of nature is to be pondered.

(An essay written in one of the earliest books on dogs, in Latin in 1570, by Johannes Caius, a Doctor of Medicine of the University of Cambridge. This English translation by Abraham Fleming, originally published by Richard Johnes, in 1576, is quoted by David and Deborah Thacker of Seashell English Setters on their website (`www.vvm.com/~dthacker`). The use of the word 'index' in both the title and the essay itself presumably relates to the Latin meaning of 'index' as 'one that informs or indicates; an informer'.)