

John Williams, Freelance Lexicographer, supported by the Open University, and other interested parties

The question of plagiarism and breach of copyright in the dictionary-making process (with particular reference to the UK)

ABSTRACT: In dictionary-making, normal definitions of plagiarism and copyright do not apply. A certain amount of similarity to previously published material is probably inevitable. The results of a questionnaire sent to dictionary project managers are used to show that there is little agreement or consistency in current practice regarding plagiarism. The inadequacies of the legal framework are considered, along with various ad hoc approaches to the problem. The paper concludes by calling for more co-operation between lexicographers on these matters.

1. Posing the problem

The question of plagiarism, or infringement of copyright, poses particular problems in dictionary publishing, distinct from those faced by the general publisher. Personnel at all levels of the dictionary-making process – from the company's lawyers, to project managers, and even low-level freelance compilers – may from time to time have to make decisions as to what constitutes, and what does not constitute, an unacceptable degree of copying from previously published dictionaries. *Prima facie* cases of plagiarism may manifest themselves in many different aspects of the dictionary text: conventions of presentation, the headword list, definition style, not to mention the actual content of particular definitions, examples, translations and illustrations.

The core of the problem resides in the fact that all lexicographers working in a given language (or between two given languages) are covering essentially the same body of knowledge. Given that this body of knowledge is itself linguistic in nature, and that it is usually incumbent upon the lexicographer to produce a text which documents, often by way of example, typical meanings and usages, it is probably inevitable that dictionaries of broadly similar scope and aims will contain comparable stretches of text that show a degree of similarity unacceptable in other areas of publishing¹. Of course, some leeway is provided by the fact that dictionaries come in different shapes and sizes, with different markets in mind, and each will carve out its own particular section of *langue* appropriate to its own purposes. However, the problem becomes particularly acute when a publisher sets out to produce a direct competitor to an already existing dictionary in terms of size, number of headwords, and general focus of coverage.

Given that most working lexicographers continually consult other dictionaries as matter of routine, it is important that all personnel involved in the dictionary-making

process should be in a position to decide, on some kind of systematic and uniform basis, what kinds of copying, if any, are acceptable, and which are not. This is desirable not only in terms of the strict legal requirements of copyright, but also on professional and ethical grounds. As a freelance lexicographer who has worked for a number of different companies, it is my suspicion that, at present, no such basis for a common practice exists among publishers of dictionaries in the UK. This paper will set out the results of my attempts to test that assumption, and it is hoped that this may provide some pointers towards establishing an agreed set of ground rules determining the extent to which dictionary publishers may legitimately “borrow” from each other’s publications without fear of litigation or professional acrimony.

2. Research material, and the difficulty of obtaining it

The original impetus behind this paper came from my experiences as a contributor to a number of lexicographical projects for UK publishing companies, on both an in-house and freelance basis. I can say without exaggeration that policies towards plagiarism range from the obsessive to the totally unconcerned. In order to support this experience with more objective evidence, I sent a questionnaire, reproduced in the Appendix, to 13 dictionary project managers working for UK publishing companies. Each of the possible answers to each question was assigned a points value, the sum of which would enable me to quantify, roughly and somewhat unscientifically, each informant’s sensitivity to plagiarism. The informants themselves were not made party to this points system².

Of 13 questionnaires sent out, replies were received in respect of 12, though this yielded a total of only 7 completed questionnaires. This is partly explained by the fact that there was, on the part of the larger companies in particular, some reluctance to divulge confidential information, or to involve themselves in what they saw as an attempt to second-guess legal opinion. In the words of one senior manager:

“I do not believe that dictionary makers or dictionary users would be helped by attempts to anticipate possible legal rulings.”

This shows, if nothing else, that there is a certain amount of concern and suspicion, at least at senior management level, over questions of copyright. Editorial staff have been, in general, rather more co-operative, and some useful material has been provided, wittingly or otherwise, by informants who did not actually complete a questionnaire.

In view of these sensitivities, and following specific requests from informants, I have decided not to identify any of my sources. This may make for frustrating reading at times, but I can guarantee that all the reported comments are genuine.

The small size of the sample must cast some doubt on the statistical validity of the findings, but it is difficult to see what other approach could have been adopted. Certainly, any attempt to base the paper round alleged actual examples of plagiarism in published dictionaries would have been folly.

3. Responses of Informants

In my experience, there appear to be two basic schools of thought on the part of dictionary project managers;

1) Any hint of plagiarism must be avoided at all costs, even if this means some loss in terms of quality and completeness. It is the task of editors to modify entries that are too similar to their equivalents in other dictionaries, even if this similarity is purely coincidental.

2) It is the duty of the lexicographer to produce dictionary entries of the highest possible quality, even if this involves close similarities with rival dictionaries. This position is well summed up by one informant who remarked, à-propos of the presentation of grammatical information: "What's the point of abandoning something good just because someone else has done it?"

This split was confirmed to some extent by the answers to questions 5 and 6, though there was a marked tendency towards the latter position. Obviously, the above statements represent extreme points of view; the degree of sensitivity to plagiarism, measured on a 100-point scale derived from the points system described in the previous section, ranged from 37 to 59. This amount of variation need not in itself be a problem; what makes it so is the fact that editorial policy towards plagiarism is generally not codified into a specific set of guidelines. There may thus be substantial differences in the manner and the extent to which individual lexicographers working on the same project make use of other dictionaries. Most projects operate in a grey area between the two extremes mentioned above; the difficulty arises when each project, or even each lexicographer, inhabits a different part of the grey area.

Some informants mentioned particular lexicographical practices which make similarities between dictionaries either more or less likely. Obviously, the increasing use of corpora reduces the probability of there being identical examples in two different dictionaries (provided that the examples are taken from the corpus without modification). On the other hand, the use of a restricted defining vocabulary limits the lexicographer's room for manoeuvre, and (s)he may have to fall back on, or may coincidentally come up with, existing definitions devised within similar limitations.

The hypothetical individual examples in Section C of the questionnaire also came in for some comment. Perhaps the most interesting issue to emerge was the difficulty of dealing with words that occur in restricted environments³; for instance, distinction was drawn between a verb such as *lack*, with its relatively unlimited set of collocates, and a verb such as *conduct*, with its much narrower range [enquiry/investigation/survey/experiment/meeting]. While the choice of identical sets of collocates would probably constitute plagiarism in the first case, this would not necessarily be so in the second case.

The striking thing about this observation, and also the point about restricted defining vocabularies, is that it is of a rather technical, lexicographical nature. If an actual dispute were to arise between companies over plagiarism, copyright legislation would be of little use unless it were sufficiently sophisticated and targeted to cope with such distinctions. Understandably perhaps, it is doubtful whether existing legal texts are able to suggest clear cut solutions to copyright problems in dictionary-making.

4. The legal framework

As far as the UK is concerned, the relevant piece of legislation is the 1988 Copyright, Designs and Patents Act, which, insofar as it affects dictionaries, does not differ in any important respects from previous legislation. For those, like myself, who are not well-

versed in legal matters, the legislation is probably best approached through standard guides and commentaries, such as those of Skone James and Dworkin. Since none of these commentators, nor indeed the legislation itself, treats dictionaries as special case, it is uncertain whether any of the following remarks would carry any legal force. However, it is clear that copyright legislation does take account of a number of concepts and issues that are most certainly of relevance to dictionary-makers.

The most important of these is probably **substantiality**, which relates to the amount or extent of copying from one work to another. The term is slightly misleading, since, as all the commentators agree, the deciding factor is **quality** rather than **quantity**. That is to say, a breach of copyright may be considered to have taken place even if only small part of one work is copied and used in another. Translated into dictionary terms, this means that a single pair of identical definitions or examples may, in theory, be grounds for legal action. Obviously, the greater the amount of similarity, the greater the probability of a successful prosecution for breach of copyright, but, as Dworkin comments, rather unhelpfully: "It is all a matter of fact and degree" (Dworkin 1989, 56)⁴.

Hence, as far as questions 36 to 49 of the questionnaire are concerned, there is no strict legal justification for considering copying to be acceptable in isolated cases, but not as a generalized practice. Only one of my informants was entirely consistent on this score. However, the alternative view might be defended on pragmatic grounds, in that isolated instances of copying will certainly be less easily detectable.

The vital question remains of what constitutes **quality** in dictionary-making. The only commentator who comes near to tackling this question is Skone James, who includes dictionaries in the category of "compilations", along with gazetteers, arithmetical tables, lists of football results, etc. Skone James recognizes that similarities between such works cannot be judged in the same way as similarities between, say, "original" works of literature. In the case of such compilations, similarity is not in itself proof of breach of copyright, though it may be used to **allege** such a breach. The onus would then be on the defendant to show that this similarity was not the result of copying; evidence might be adduced from textual detail, or from the circumstances in which the text was produced⁵. A successful defence might be that the similarity was due to coincidence or "subconscious copying", though the law is undecided on this latter point⁶.

The key criterion throughout appears to be whether or not the copier has unfairly appropriated the labour, energies and resources of the compiler of the source material:

"...the principle is that if one person has, with considerable labour, compiled a work from various sources which he has digested and arranged, then a defendant who instead of taking the pains of researching the common sources and obtaining his subject matter from them, simply makes use of the other's labour and adopts his arrangement, perhaps with only slight variations, thus saving himself the pains and labour which the other used, this will be an illegitimate use." (Skone James 1991, 180)

The precise implications of these legal principles for working lexicographers are not immediately clear, since none of the commentators addresses problems that are specific to dictionaries such as words that occur in restricted environments, words or set phrases with a single standard translation, etc. However, one might hazard a few tentative conclusions. Firstly, that definitions, examples, translations, etc. which "spring readily to mind", casual phrases used to fill out examples⁷, and standard conventions of presentation would probably not be subject to copyright protection, whereas, for example, more

“difficult” definitions and translations, and more innovative types of presentation, probably would be⁸. Secondly, as the above quotation makes clear, that minor variations in an otherwise plagiarized text could not be used to refute allegations of breach of copyright⁹. Thirdly, that the practice routinely adopted by working lexicographers of having the dictionaries of our competitors open in front of us makes it more likely that a prosecution for breach of copyright would be successful.

The above remarks, however, probably beg more questions than they answer; the fact that dictionaries do not figure at all prominently in legal texts concerned with copyright is a serious handicap to anyone wishing to formulate a legally watertight set of guidelines for dictionary-makers.

5. Ad hoc approaches to plagiarism

In the absence of any sound legal information, project managers have sometimes devised their own, often rather idiosyncratic, tests for plagiarism. One clever little device known to have been used occasionally in dictionary projects is the **bugword**, familiar to about half my informants. A bugword is non-existent word deliberately included in a dictionary, so that it can be used to support allegations of copying if it reappears in the dictionaries of other publishers¹⁰. None of my informants admits to using these, though I did receive the following comment:

“One of the editors tried to put some [bugwords] in but I thought that was being too clever by half and had them taken out before we went to press.”

The consensus seems to be that bugwords are rather unsound, if amusing, lexicographical device, and so probably best avoided.

More deserving of serious attention are the various rules of thumb designed to prevent too great a degree of similarity between dictionaries. The questionnaire tested informants’ familiarity with the so-called **ten-word rule**, which outlaws sequences of ten words or more that are identical to sequences at equivalent places in other dictionaries. Only one informant claims to apply some form of this rule, and (s)he interprets it, untypically, as requiring that an acknowledgment be given in cases where such identical sequences occur. The ten-word rule is obviously inadequate to deal with certain lexicographical items such as long proverbs, and set phrases plus standard translation. Conversely, it is possible that, in law, much shorter sequences of words could constitute clear cases of plagiarism. One informant reports having operated an even more draconian **four-word rule**.

Other rules of thumb cited include the avoidance of two consecutive columns of headwords corresponding exactly to those of another dictionary, and a general requirement for compilers to draw up draft entry for each item before consulting other dictionaries.

None of these rules of thumb can be said to flow directly from existing legal requirements. However, they do all have the advantage, when applied, of reducing the risk of a breach of copyright. Their value would be significantly greater if they were uniformly adopted by all UK dictionary publishers.

6. The way forward

With such a small sample of responses, it is difficult to find in the questionnaire any basis for formulating a specifically lexicographical definition of plagiarism that would be acceptable to most dictionary publishers. The responses to Questions 39 and 40 may perhaps provide a small pointer; it seems that, at least as far as the definition of countable common nouns is concerned, it is generally acceptable to retain an existing definition as long as a minimum of two lexical elements of it are altered, even if this constitutes fewer than half the lexical elements of the definition (as in the case of *lever*). A similar rule might conceivably be applied to examples. In contrast, informants are less convinced by superficial attempts to avoid direct copying which simply alter some of the **function words** in the definition or example.

Again, it must be stressed that the above suggestions have no foundation in copyright law. However, in the absence of any "lexicographical legislation" or test cases, it seems preferable that the practice of dictionary-makers should be based on agreement, rather than legal technicalities. In this light, it is worth noting the fact that only a minority of informants report having discussed the problems of plagiarism with senior editorial staff from other companies. It is hoped that this paper, and the discussion it arouses, will be a step on the road to remedying that situation.

Endnotes

1. Though two of my informants would appear to disagree with this point of view.
2. I have not given details of this points system in the Appendix, but will gladly do so to anyone who expresses an interest.
3. I had intended Question 46 (**navigate**) to be an example of this, though only one informant explicitly picked it out as such.
4. Skone James concurs, rather more wordily: "In short, the question of substantiality is a matter of degree in each case and will be considered having regard to all the circumstances. Generally, it is not useful to refer to particular decisions as to the quantity taken." (1991, 175)
5. "But a bare assertion that the plaintiff's work was not copied, without any explanation by the defendant as to how and when he worked or how long it took him may well not be enough to rebut the inference of copying. Whereas, with a strictly original work, any identity of phrase is sufficient evidence of copying, with many compilations it is only from external evidence, or from a minute examination of textual errors, that an infringement can be established." (Skone James 1991, 176-177)
6. See Skone James 1991, 171.
7. Though they could probably be used as supporting evidence in cases of more "substantial" copying.
8. There is a general maxim in publishing, that "what is worth copying is **prima facie** worth protecting", though Skone James adds an important rider to this (1991, 175-176).
9. Though this probably depends on the level of ingenuity involved.
10. I first heard of this practice in connection with an American corpus of children's playground slang. However, I am unable to give a precise reference.

Bibliography

- BURCHFIELD, Robert (1984): "Dictionaries New & Old: Who plagiarizes Whom, Why & When?" In: *Encounter* Sept./Oct. 1984. (reprinted in *Unlocking the English Language*. Faber & Faber, London, 1989)
- BUTCHER, Judith (1987): *Copy-Editing*. Cambridge University Press, Cambridge.
- DWORKIN, Gerald (1989): *Blackstone's Guide to the Copyright, Designs and Patents Act 1988*. Blackstone Press Limited, London.
- MICHAELS, Amanda L. (1992): "British Copyright Law". In: *the Writers' and Artists' Yearbook 1992*. A & C Black, London.
- SCARLES, Christopher (1980): *Copyright*. Cambridge University Press, Cambridge.
- SKONE JAMES, E.P. (et al.) (1991): *Copinger and Skone James on Copyright*. Sweet and Maxwell, London.

Thanks also to Don Emerson, S.B., Robert Burchfield, and to all those who completed the questionnaire.

20. Do you instruct editorial staff to check the text produced by other editors for plagiarism?
 YES: 1 NO: 6 DON'T KNOW/DEPENDS: 0
21. Would you ever seek to dismiss an employee, or dispense with the services of a freelance, for persistently copying from other dictionaries?
 YES: 6 NO: 0 DON'T KNOW/DEPENDS: 1
22. Has your company ever instituted legal proceedings for breach of copyright against the publishers of another dictionary?
 YES: 0 NO: 7 DON'T KNOW/DEPENDS: 0
23. Have the publishers of another dictionary ever instituted legal proceedings against your company for breach of copyright?
 YES: 0 NO: 7 DON'T KNOW/DEPENDS: 0
24. Does your company regularly check dictionaries produced by rival companies for evidence of plagiarism?
 YES: 3 NO: 4 DON'T KNOW/DEPENDS: 0
25. Are you familiar with the "ten word rule"?
 YES: 2 NO: 5 DON'T KNOW/DEPENDS: 0
26. Do you apply the "ten word rule" in your projects?
 YES: 1 NO: 5 DON'T KNOW/DEPENDS: 1
27. Do you apply any other rules of thumb in your projects?
 YES: 2 NO: 3 DON'T KNOW/DEPENDS: 2
28. Are you familiar with the term "bugword"?¹
 YES: 3 NO: 4 DON'T KNOW/DEPENDS: 1
29. Do you use "bugwords" in your project?¹
 YES: 0 NO: 6 DON'T KNOW/DEPENDS: 2
30. Who does the responsibility for avoiding plagiarism lie with primarily?²
- | | |
|-----------------------------------|---|
| a) The compiler of the entry | 7 |
| b) The editor checking the entry | 1 |
| c) Senior editorial staff | 1 |
| d) The project manager | 0 |
| e) The company's legal department | 0 |
| f) Other – please specify | 0 |
| g) Don't know | 0 |
31. What system of payment do you use for freelancers?¹
- | | |
|--|---|
| a) By the hour, the number of hours worked being assessed by the freelance him/herself | 4 |
| b) A predetermined payment for each batch of text | 3 |
| c) Both (a) and (b), or a combination of the two | 2 |
| d) Other – please specify | 0 |
| e) Don't use freelancers | 0 |

Section C: Examples of plagiarism³

Imagine that X and Y are two dictionaries competing in the same market, the compilation stage of Y not beginning until after X was published. In which of the following cases would you consider that Y exhibits an unacceptable degree of similarity with respect to X.

(YES = unacceptable; NO = acceptable)

32. Y adopts identical conventions of typeface, punctuation and abbreviation to X, but in all other respects the dictionaries are substantially different.
 YES: 3 NO: 3 DON'T KNOW/DEPENDS: 1
33. Y has exactly the same headwords as X, but the content of the entries is substantially different.
 YES: 5 NO: 1 DON'T KNOW/DEPENDS: 1
34. Y presents grammatical and syntactical information in identical fashion to X, but in all other respects the dictionaries are substantially different.
 YES: 3 NO: 2 DON'T KNOW/DEPENDS: 2
35. Y is identical to X in all the respects described above, but the content of the entries is substantially different.
 YES: 5 NO: 0 DON'T KNOW/DEPENDS: 2
36. X: **laburnum** (n) small ornamental tree with hanging clusters of yellow flowers
 Y: **laburnum** (n) small ornamental tree which has hanging clusters of yellow flowers
 – as an isolated instance YES: 1 NO: 6 DON'T KNOW: 0
 – as a typical instance YES: 7 NO: 0 DON'T KNOW: 0

37. X: **laburnum** (n) small ornamental tree which has hanging clusters of yellow flowers
 Y: **laburnum** (n) small ornamental tree with clusters of yellow flowers that hang down
 - as an isolated instance YES: 1 NO: 6 DON'T KNOW: 0
 - as a typical instance YES: 3 NO: 2 DON'T KNOW: 2
38. X: **ladykiller** (n) man with the reputation of being very charming and popular with women
 Y: **ladykiller** (n) man considered to be very charming and popular with women
 - as an isolated instance YES: 1 NO: 6 DON'T KNOW: 0
 - as a typical instance YES: 5 NO: 2 DON'T KNOW: 0
39. X: **ladykiller** (n) man with the reputation of being very popular and successful with women
 Y: **ladykiller** (n) man considered to be very charming and popular with women
 - as an isolated instance YES: 1 NO: 6 DON'T KNOW: 0
 - as a typical instance YES: 1 NO: 5 DON'T KNOW: 1
40. X: **lever** (n) bar or other device turning on a fixed point which lifts or opens something with one end when pressure is applied to the other end
 Y: **lever** (n) rod, handle or similar device turning on a fixed point which raises or opens something with one end when force is applied to the other end
 - as an isolated instance YES: 0 NO: 7 DON'T KNOW: 0
 - as a typical instance YES: 2 NO: 5 DON'T KNOW: 0
41. X: **lag** (vi)... (Example) he was lagging behind the others..
 Y: **lag** (vi)... (Example) she was lagging behind the others..
 - as an isolated instance YES: 1 NO: 6 DON'T KNOW: 0
 - as a typical instance YES: 4 NO: 2 DON'T KNOW: 1
42. X: **lame** (adj)... (Example) the horse is lame in one leg..
 Y: **lame** (adj)... (Example) the horse was lame in one leg..
 - as an isolated instance YES: 2 NO: 4 DON'T KNOW: 1
 - as a typical instance YES: 6 NO: 1 DON'T KNOW: 0
43. X: **label** (vt)... (Example) every packet must be clearly labelled..
 Y: **label** (vt)... (Example) every bottle must be clearly labelled..
 - as an isolated instance YES: 3 NO: 4 DON'T KNOW: 0
 - as a typical instance YES: 7 NO: 0 DON'T KNOW: 0
44. X: **label** (vt)... (Example) every bottle must be clearly labelled..
 Y: **label** (vt)... (Example) the bottle was clearly labelled..
 - as an isolated instance YES: 0 NO: 7 DON'T KNOW: 0
 - as a typical instance YES: 4 NO: 2 DON'T KNOW: 1
45. X: **label** (vt)... (Examples) every packet must be clearly labelled... the bottle was not labelled..
 Y: **label** (vt)... (Examples) every bottle must be clearly labelled... the packet was not labelled..
 - as an isolated instance YES: 6 NO: 1 DON'T KNOW: 0
 - as a typical instance YES: 7 NO: 0 DON'T KNOW: 0
46. X: **navigate** (vi)... (Example) you drive, I'll navigate..
 Y: **navigate** (vi)... (Example) I'll drive, you navigate..
 - as an isolated instance YES: 2 NO: 5 DON'T KNOW: 0
 - as a typical instance YES: 5 NO: 1 DON'T KNOW: 1
47. X: **lament** (vt) (Translation) pleurer, regretter, se lamenter sur
 Y: **lament** (vt) (Translation) se lamenter sur, pleurer, regretter
 - as an isolated instance YES: 1 NO: 2 DON'T KNOW: 4
 - as a typical instance YES: 1 NO: 2 DON'T KNOW: 4
48. X: **lack** (vt) (confidence, strength, friends, interest) (Translation) manquer de
 Y: **lack** (vt) (friends, confidence, interest, strength) (Translation) manquer de
 - as an isolated instance YES: 5 NO: 1 DON'T KNOW: 1
 - as a typical instance YES: 6 NO: 0 DON'T KNOW: 1
49. X: **lack** (vt) (strength, friends, interest) (Translation) manquer de
 Y: **lack** (vt) (friends, strength, confidence) (Translation) manquer de
 - as an isolated instance YES: 2 NO: 4 DON'T KNOW: 1
 - as a typical instance YES: 4 NO: 2 DON'T KNOW: 1

Endnotes

1. Total exceeds 7 because answers can reasonably be deduced from material made available by informants who did not actually complete a questionnaire.
2. Total exceeds 7 because one informant gave more than one answer.
3. Any resemblance between particular examples in this section and actual published dictionary entries should not be interpreted as alleging any acts of plagiarism on the part of the publishers in question; that is not the purpose of this paper.