Making Room for the Compound Nouns
in Small Monolingual English Dictionaries

Dr. Della Thompson

This paper makes a plea for a fuller coverage of compound nouns in small English dictionaries. This is not to say, however, that these words have been neglected over the years. On the contrary, their treatment has expanded, at least in proportion to the number of headwords included: for example, the first editions of the Concise Oxford Dictionary (1911) and the Pocket Oxford Dictionary (1924) contained 23 and 13 compounds of head respectively, compared with nearly twice as many (45 and 22) in the seventh editions. Whether this increase is proportional to the increase in the number of new compounds coined would be difficult to determine, but it still shows an awareness among lexicographers of their importance.

However, when I came to revise the Pocket Oxford Dictionary, it seemed to me that one of its weaknesses lay in its coverage of compound nouns. For example, under head, I found the following words missing: headband, headbutt, headcount, headhunter, headhunting, headrest, headscarf, headset, and head teacher, all of which seemed worthy of inclusion in a dictionary the size of the POD. When I looked at other small dictionaries for English native speakers, I found similar omissions: Collins Pocket Dictionary also left out all these words, except headhunter, headhunting, and head teacher, while including head-banger, not in the POD; Chambers Pocket Dictionary revealed further discrepancies: it listed headband, headhunting, and headset but left out headgear, headlamp, headshrinker, headstall, and headwaters, all of which appear in the other two dictionaries, and gave headpiece as an extra word not appearing in the other two works: the Longman Pocket English Dictionary was different again, listing headband, head-hunting, headrest, and headset, while omitting headbutt, headcount, headgear, headhunt, headlamp, headscarf, and headwaters. These discrepancies can be shown more clearly in tabular form:

<table>
<thead>
<tr>
<th>Coverage of Compounds of head in Pocket English Dictionaries</th>
</tr>
</thead>
<tbody>
<tr>
<td>POD7</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>headband</td>
</tr>
<tr>
<td>head-banger</td>
</tr>
<tr>
<td>headbutt</td>
</tr>
<tr>
<td>headcount</td>
</tr>
<tr>
<td>headgear</td>
</tr>
<tr>
<td>headhunter</td>
</tr>
<tr>
<td>headhunting</td>
</tr>
<tr>
<td>headlamp</td>
</tr>
<tr>
<td>headman</td>
</tr>
<tr>
<td>headpiece</td>
</tr>
<tr>
<td>headrest</td>
</tr>
</tbody>
</table>
Coverage of Compounds of *head* in Pocket English Dictionaries

<table>
<thead>
<tr>
<th></th>
<th>POD7</th>
<th>Collins Pocket</th>
<th>Chambers Pocket</th>
<th>Longman Pocket</th>
</tr>
</thead>
<tbody>
<tr>
<td>headscarf</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>headset</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>headshrinker</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>headstall</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>head teacher</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>headwaters</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>headword</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fifteen compound nouns on the list appear inconsistently in the four dictionaries looked at, while the remaining three appear in none of them. This compares with fifteen compounds of *head* not on the table but present in all four dictionaries: *headache, headboard, headdress, headland, headlight, headline, headmaster, headmistress, headphone, headquarters, headroom, head start, headstone,* and *headwind.* Only three words, *headbanger, headrest,* and *head teacher,* appear only in one dictionary, while nine appear in two and three in three of the four works.

I would argue that all these words, with the exception of *headpiece,* deserve to be included in the next edition of the *POD.*

When I compared these findings with a range of non-compound headwords, I did not find the same discrepancy in coverage between comparable dictionaries. In the range *hallucinate* to *hammer,* occupying one column of text, the differences between *POD7* and the *Collins Pocket* were minor: the *POD* included the headwords *halm* (variant of *haulm*), *halves* (plural of *half*) and *Hamitic,* while Collins included *halting* and *hamba.* The most striking difference between the two dictionaries in this range of text was again in the coverage of compound nouns: Collins included *hallway* and *halterneck,* both absent from the *POD.*

The coverage of compounds seems, therefore, to be an area of dictionary compilation which, like the coverage of minor senses, idioms, phrasal verbs, derivatives, and variants, is left largely to the discretion of the individual editor: if we were to examine any two dictionaries of similar size and scope we would find many variations in these other areas too. Many compound nouns, it seems, fall into the category of «marginal» vocabulary.

In order to try and discover whether different dictionaries had different policies for the inclusion of compound nouns, I studied the introductions of many of them, but found that most make no remarks at all about compounds. Two exceptions are the *Oxford English Dictionary* and the *Concise Oxford Dictionary.* The *OED* in its first edition acknowledges the difficulty of determining which «combinations» are to be dealt with by the lexicographer, and divides them into three groups, one of undefined nested items, one of defined nested items, and one of items with full headword status. The undefined items are considered to be semantically «transparent».* The *COD* does not mention compounds in its introduction until the sixth edition (1976) when it
merely states: «Limitations of space make it impossible to include every possible and legitimate derivative (e.g. compounds of obvious meaning such as boiler-room...»). The seventh edition repeats this statement but the eighth wisely falls silent again on the subject.

The COD (until the eighth edition), like the OED before it and every other dictionary I have consulted, did not consider boiler-room worthy of attention, because its meaning was «obvious». However, both dictionaries gave engine-room which was arguably no less obvious. One could also argue that bedroom and bathroom are also fairly obvious and yet it would be almost impossible to find a dictionary that did not contain them. I would suggest, therefore, that their inclusion is on the grounds of currency, and currency alone, rather than on the grounds of, say, productiveness or spelling. This brings me to what I think to be the most important reason for including a compound in a dictionary, which is the user’s expectation that it should be there. A common everyday word such as bathroom must be in a dictionary, just as kitchen must be, whether it is semantically transparent or not. Surely, words such as headband, headrest, and headscarf also deserve a place, at least in a dictionary the size of the POD, for the same reason.

We all know that one of the main uses to which a dictionary is put is to check spelling, and that in connection with this hyphenation practice is a matter of interest and importance to users of all kinds. This is another good reason for including as many compounds as possible in dictionaries, especially in English dictionaries. In addition, the need to know simply whether a word «exists» or is a «real word» is often paramount among players of word-games such as Scrabble, and therefore dictionaries need to be explicit rather than implicit in the information they give.

What, then, are the ways in which space can be found to accommodate these perhaps underrepresented vocabulary items in a climate which is encouraging us to use up more space than ever before by denesting, using definite and indefinite articles in definitions, providing usage notes and grammatical information, and generally being more user-friendly? In my abstract I suggested three ways in which space might be saved, the first two of which turned out to be not such a good idea when examined more closely.

Firstly, I recommended the introduction into our smaller dictionaries of lists of undefined compounds of the boiler-room type, as was done in the OED, on the grounds that we consider this to be adequate treatment in the POD of in-, un-, re- and over- words, while Collins... Pocket uses it for words in un-, re-, non-, and under-. This method of saving space no longer seems advisable. Even the OED, which claims to list undefined compounds, in fact gives some guidance to their meaning. For example, before the list which includes hair-clip, hair-conditioner, hair-cream, and hair-spray, it gives what amounts to a definition which serves for the whole list: «for or for the use of the hair», and the list which includes hair-dryer and hair-straightening is prefaced by the definition «objective and objective genitive». Secondly, in the interests of user-friendliness, in-, un-, re-, and over- words will probably no longer be placed in lists at the bottom of the page in the next edition of the POD but will be individually defined as in the COD. It would, therefore, be a retrograde step to introduce new lists of undefined words. Thirdly, there are very few compound nouns whose precise meaning can be deduced from their elements. In this sense, they are unlike many words with prefixes such as un- and re-, whose meaning is clear from their elements.

My second suggestion for saving space was to omit some of the more common
idioms that we give at the moment, e.g. *keep (or lose) one's head*, on the grounds that they are generally understood by native speakers, do not present spelling problems, and are probably not automatically expected to be there. This suggestion fails really on the first of these reasons, that they are very common and therefore readily understood by native speakers. It is extremely difficult to make accurate assumptions as to what the user knows or does not know, and, besides, what we include in dictionaries is, at least at the moment, based on currency rather than rarity.

My third suggestion was that space could be saved by making definitions and examples shorter. On further examination, this proved to be the most fruitful method of shortening dictionary entries, with definitions turning out to be far more fruitful than examples. In order to see just how much space could be saved, I took a section of about $1^{1/2}$ pages of edited text in letter «B», from *back* to *backyard*, and set about reducing it as much as I could. Having done this, I analysed the different types of shortening which had taken place. Nine different types emerged, which are as follows:

(i) The use of shorter simpler words or a single word in place of a phrase, e.g.:

- **back v. 1a**  
  *help with* moral or financial support ⇒ *give* moral or financial support to

- **back down**  
  withdraw one's claim or *point of view* ⇒ withdraw one's claim or *argument.*

- **back number**  
  issue of a periodical *earlier than the current one* ⇒ *out-of-date* issue of a periodical

- **backgammon**  
  *game played on a board* with pieces moved according to throws of the dice ⇒ *board-game* with pieces moved according to throws of the dice

- **backwash**  
  receding waves *created* by the motion of a ship, etc. ⇒ receding waves *made* by a ship etc.

(ii) The deletion of superfluous words, e.g.:

- **back v. 1b**  
  bet on *the success of* (a horse etc.) ⇒ bet on (a horse etc.)

- **backwash**  
  receding waves *created by the motion of* a ship etc. ⇒ receding waves *made by* a ship etc.

(iii) The deletion of uncommon or suspect elements, e.g.:

- **back n. 3a**  
  *less active or visible or important part* ⇒ *less visible or important part*
(iv) The deletion of superfluous phrases such as «applied to», «denoting», etc., e.g.:

**back-to-nature**  
_applied to a movement or enthusiast for reversion to a simpler way of life_  

seeking a simpler way of life.

**backstreet**  
_denoting illicit or illegal activity_  

illicit; illegal

(v) The deletion of synonyms or near synonyms, e.g.:

**back**  

_side or part normally away from the spectator..._  

part normally away from the spectator...

**get (or put) a person’s back up**  

_annoy or anger a person_  

annoy a person

**back seat**  

_less prominent or important position or status_  

less prominent or important position

(vi) The deletion of tautologous words, e.g.:

**backlog**  

_arrears of uncompleted work_  

arrears of work.

(vii) Making overlong definitions shorter, even if some of the sense is lost, e.g.:

**backlist**  

_publisher’s list of books published before the current season and still in print_  

publisher’s list of books in print.

(viii) Avoiding redefining elements of simple words defined elsewhere, e.g.:

**backache**  

_(usu. prolonged) pain in the back_  

ache in the back.
(ix) The use of a single definition for transitive and intransitive senses of the verb where the form of the defining verb is identical for both, e.g.:

**back v. 2**  move or cause to move backwards ⇒ move backwards.

The result of this space-saving exercise was extremely gratifying. In a total of 201 lines, 15 lines were saved, or 7.5%. In a book the size of the *POD* this would amount to 70 pages if carried out on the whole text. It was also interesting to see how savings were divided between long and short entries. I had imagined that proportionally more space would be saved by cutting longer entries rather than shorter entries since in shorter entries savings are lost if they are less than the length of the final line of the entry. In the larger entries they would, I thought, be cumulative and could only be lost once, in the last line. However, the opposite turned out to be the case, obviously because a line can often be saved by deleting less than a line and this opportunity arises more frequently in a run of short entries. In the long entry, *back*, 188 characters were deleted, saving 5 lines (i.e. one line was saved for each 37.6 characters deleted), while in the range *backache-backyard* 244 characters were deleted, saving 10 lines (i.e. one line was saved for each 24.4 characters deleted). Assuming that the section taken was representative, it seems that half as much text again has to be cut from a long entry in order to save the same amount of space as from cutting shorter entries.

Two other benefits of such an exercise are that definitions are improved overall and errors may be spotted and rectified.

The main disadvantage of such an exercise is that it is very time-consuming. Not only does it take time to think up briefer ways of expressing things, it takes even longer to check in the *OED*, quotation files, corpora, or simply in one’s head, that one is not omitting something important when removing what appears to be superfluous material. The other danger is that definitions could become less user-friendly by being pared down to a succinct minimum.

Having saved perhaps 7.5% of the volume of the dictionary in this way, the compounds that might well be considered for inclusion in this section are *backbend*, *backflip*, *backlighting*, *backrub*, *backscratcher*, *backswing*, and *backing track*. Of these, only *backscratcher* can be found even in the *COD*.

There is one further area, not mentioned in the abstract to this paper, in which space might well be saved in small monolingual English dictionaries. This is pronunciation. I believe that few native speakers are aware that a dictionary gives guidance on pronunciation and still fewer use their dictionary for this purpose. Even if they did, most would not understand the IPA system (now adopted by most dictionaries) which is known by probably only a few thousand people in the country, mainly language graduates and teachers of English as a foreign language. Other pronunciation systems such as that used by Longman (alongside the IPA) tend to be difficult or incomprehensible in their own way, e.g. the use of «ie» for the sound \æ\, «uh» for \ə\, and «zh» for \3\. Chambers is also difficult with «ä» representing \a\, «ö» \ő\, and «ü» \u\. All pronunciations could, I believe, be left out of dictionaries the size of the *POD* downwards with few people noticing it, let alone being inconvenienced.

To sum up, the following points can be made:
(i) There should be a fuller coverage of compound nouns in our dictionaries because they are expected to be there, they are not always as transparent as we might imagine, and because they cause spelling difficulties. Even if we cannot include as many as we would like, it may be useful to examine all the possible entries in order to decide which ones are the most important. Perhaps there could be a greater consensus than is at present the case as to which of the words in the table of «head» words should be included in a small dictionary. The availability of large corpora from which concordances can readily be made should make this task much simpler and more efficient in the future.

(ii) If we want to maintain our policies of user-friendliness, the solution to the space problem is not to revert to the old methods of nesting, using the swung dash, etc., but to look at ways of saving space in definitions. We might also consider omitting pronunciations from our smaller English dictionaries for native speakers.