3L: Language, Linguistics, Literature® The Southeast Asian Journal of English Language Studies
Vol 28(3), September 2022 http://doi.org/10.17576/3L-2022-2803-05

Where Should English Dictionaries Place Multi-Word Expressions? A Study of User Consultation Behaviour

ROBERT LEW
Faculty of English,
Adam Mickiewicz University, Poznań, Poland
rlew@amu.edu.pl

EWA TOMCZAK
Faculty of English,
Adam Mickiewicz University, Poznań, Poland

ABSTRACT

Multi-word expressions (MWEs) are a known challenge to lexicographers and language learners alike. Dictionary articles are traditionally organized around single words, so lexicographers are forced to choose the headword under which to include a given MWE. When users encounter a problematic MWE in text, they also face the choice of the component word under which to look it up. The present study attempts to find out what factors can predict such lexical choices so that lexicographers can position the information in the very entry that dictionary users will choose to consult. Based on previous findings, four factors are selected as potentially predictive: lexical frequency, part of speech, word position, and word length. A list of thirty-seven English sentences with MWE's was presented to 155 Polish students at high school and university levels, asking them to indicate which word they would choose to look up in a dictionary. Our analysis suggests that these choices are best explained in terms of lexical frequency, with a preference for less frequent items. Empirical results are then compared with actual lexicographic practice for these very MWE's in five major English learners' dictionaries, indicating there is much room for improvement, and that such improvement is easily achieved using our findings. Finally, it is argued that the issue of MWE lemmatization is relevant for print and digital dictionaries alike.

Keywords: multi-word expression; MWE; dictionary; dictionary consultation; dictionary access; lexicography; English language

INTRODUCTION

Multi-word expressions present a special challenge to language learners and lexicographers alike. They are indispensable yet difficult to language learners (Tomczak & Lew, 2019), as they match a sequence of forms (several words) with a more or less unitary meaning; this runs counter to the naive view of the lexicon, under which words are 'containers' of meaning. Multi-word expressions (MWE's) are also difficult to dictionary makers (Atkins & Rundell, 2008), as dictionaries—by their very nature—also reflect the container view of meaning, being prototypically organized by clustering information into entries headed by a single word (appropriately known as a *headword*). English MWE's are of particular importance, given the unique status of English as a *de facto lingua franca* in many parts of the world, and many multi-lingual contexts.

A multi-word expression in a dictionary will typically be nested under (and discoverable via) one of its component words (Heuberger, 2020). The crucial decision is *which* of the component words should be selected as the keyword for the expression? To take a concrete example, should the expression *wash one's dirty linen in public* be entered under the verb *wash*, the noun *linen*, or

maybe under *dirty*, or else *public*? It is generally recognized that function words (such as the preposition in our example), or special 'placeholder' words such as *one's* or *your* (to be filled in by another word in actual use) would *not* be reasonable choices of the headword under which to enter the expression. Beyond these useful but limited generalities, however, it is far less obvious which of the remaining content words *would* be the best choice.

In dictionaries for learners of English, a fairly common policy is to list the multi-word expression *under the first content word* (in our example that would be the verb *wash*). A content word is a lexically rich word such as a noun, verb, adjective, or adverb, as opposed to grammatical (or function) words which tend to be semantically empty and carry grammatical functions or relations, such as prepositions, pronouns, determinatives. This policy will sometimes be stated overtly in the front matter of the dictionary: in the *Macmillan English Dictionary*, users are told to '[l]ook for fixed expressions at the entry for the first main word in the expression' (Rundell, 2007). However, even this stated policy is not always adhered to (Rundell, 2015, personal email communication); and more often than not, a statement of policy is not present at all.

In newly-created (born-digital) dictionaries, the current best practice appears to be to treat multi-word expressions as separate objects. However, recent research into user consultation strategies suggests that users largely continue to employ the traditional strategy of accessing multi-word expressions via a component word (Farina, Vrbinc, & Vrbinc, 2019; Knežević, Halupka-Rešetar, Miškeljin, & Milić, 2021; Wojciechowska, 2020). The problem is compounded by the fact that some multi-word expressions exhibit variability in the form (Dalpanagioti, 2021; Parizoska & Petrovic, 2020). The same problem is articulated in a recent study of the lexicographic treatment of multi-word expressions in academic discourse (Chen & Zhao, 2022).

PREVIOUS RESEARCH INTO MULTI-WORD LOOK-UP STRATEGIES

PREVIOUS FINDINGS

There appear to have been half a dozen studies looking at dictionary user strategies when consulting multi-word expressions, all focused on monolingual learners' dictionaries (MLD's): Béjoint (1981), Tono (1987), Bogaards (1990, 1991, 1992), Atkins and Knowles (1990), Atkins and Varantola (1998) (the last two are based on the same data), and Lew (2012). In what follows, we report the principal findings relevant to the present study.

Béjoint (1981) presented French students of English with a list of eight English expressions (*artificial insemination*, *boil down to*, *false alarm*, *magnetic tape*, *come down with*, *lose sight of*, *rid of*, *fountain pen*). He found a preference for head nouns in nominal compounds (note that in French head nouns tend to be the phrase-initial elements). When verbs were present in an MWE, the students preferred to go with that verb rather than an adverb or preposition (except in *lose sight of*, which exhibited a slight preference for the noun).

Tono (1987) looked into the choices of 129 Japanese learners of English across 62 idioms representing several syntactic patterns. First, content words were preferred over function words. Second, participants tended to be attracted to less familiar words. Third, Tono found a preference for words with more restricted combinability. The article is not explicit as to how exactly familiarity and combinability were measured, but it appears that they were assessed impressionistically post-hoc, which means these generalizations should be treated with caution.

Bogaards (1990) used 52 English multi-word expressions with two groups of speakers having either French or Dutch as their L1 and found quite consistent consultation behaviour within each language group, but varying across the two L1 backgrounds. In the French-speaking sample, participants seemed to prefer looking up words of lower frequency, but also syntactically superordinate terms (which, again, in French tend to be initial). Dutch speakers, by contrast, exhibited a word-class-based preference for nouns, then adjectives, and finally verbs. Two follow-up studies (Bogaards, 1991, 1992) further confirmed the frequency effect in the choices of native speakers of French.

The two reports on the *EURALEX/AILA Research Project on Dictionary Use* (Atkins & Knowles, 1990; Atkins & Varantola, 1998) do not tell us much about MWE look-up patterns because that part of the study only employed four items, one of which was discarded as problematic. Worthy of note is the finding that the look-up behaviour did not vary by the participant's L1 (French, Italian, German, or Spanish). The authors also report that the participants' preferences were all too often at odds with the lemmatizing choices of the leading monolingual learners' dictionaries (though one should be careful with a generalization based on a mere three items).

Lew (2012) looked at the role of frequency, part of speech, and word position within the MWE. Quite in line with findings from earlier studies, he found frequency to have the strongest predictive power in terms of explaining the choices of Polish-speaking participants. Part of speech played only a minor role beyond the observation that function words are for the most part ignored, but, again, slight as it was, the preference seemed to be as follows: nouns, then adjectives, then verbs, then adverbs and anything else. Finally, the position of the word within the multi-word was not found to matter.

Generalizing across the studies reviewed above, it appears that lexical frequency is a major factor that can predict under which lemma a dictionary user will seek information on the meaning of multi-word expressions in their foreign language. The role of syntactic class (or part of speech), structure, and word order remains uncertain, although it is clear that users tend not to look under function words. This last point is what MLD's get right; beyond that, however, user choices appear to diverge from lexicographic practice, with lexicographic treatment of MWE's not always given under headwords that users choose to consult.

METHODOLOGICAL PROBLEMS

The studies reviewed above exhibit some methodological problems. Samples are sometimes (though not always) small, making generalization difficult. Data analysis tends to be incomplete or impressionistic. Perhaps most importantly, all studies so far have questionable ecological validity in how they present the MWE's. In all these studies, the multi-word expressions are presented (1) taken out of context (rather than embedded in a sentence); and (2) with explicit information about their nature and extent. In effect, instructions given to participants might be loosely expressed as follows: 'here is a multi-word expression; under which component word would you look it up?'. Such testing conditions differ markedly from how MWE's would typically be encountered in real-life situations: multi-word expressions are not normally explicitly flagged in a text as such. Instead, readers may become aware of a comprehension problem but without realizing that what they are dealing with is a multi-word expression; and even if they do suspect this, they do not necessarily know the textual extent of the MWE.

THE STUDY

RESEARCH AIM AND DESIGN

Given the published findings so far, we wanted to check to what degree the three 'usual suspects' (i.e., lexical frequency; part of speech; and position in the MWE) played a role in students' choices of words to look up. In addition, we brought in one new factor that had not heretofore been considered: word length in letters. Word length in general tends to be inversely related to lexical frequency: this means that frequent words tend to be shorter, while rare words tend to be longer, by the principle of economy (Zipf, 1949). Beyond this, however, longer words may presumably be more salient in the reading process by virtue of occupying more visual space, and for this reason, word length might perhaps turn out to be a better predictor of user look-up than frequency. Here, then, are the four factors considered, with predicted directions of effect on look-up probability:

- lexical frequency bands after *Macmillan English Dictionary Online*, which indicates the bands with stars next to entries (rare > frequent);
- POS (N > Adj > V > Adv > ...);
- position in the MWE (initial > non-initial);
- word length in letters (longer > shorter).

PARTICIPANTS

Two cohorts of Polish learners of English served as participants in the study. A more advanced group of 78 students (henceforth: U) attended the final (third) year of an English-major BA programme. A lower-proficiency group were 78 final-year high-school students (henceforth: HS). Data from one HS subject were excluded from analysis, as she had only underlined anything in two items, yielding usable data from 77 high-school students.

TEST ITEMS

Thirty-seven MWE's, mostly idioms as per Granger and Paquot (2008), were presented in six different versions with randomized order of items, to minimize any order effects. For randomization, we used true random numbers as generated by https://www.random.org.

As stated in the literature overview above, all previous studies into multi-word look-up strategies presented multi-words in isolation and identified them explicitly in the testing material as MWE's. Such design revealed to the tested participants where exactly the items of interest were in the test materials, which is not what would normally happen in real-life situations. To deal with this validity issue, in the present study the target multi-word expressions to be looked up were presented as embedded in sentential contexts, and without explicitly identifying them as specific chunks (i.e., their extent was not marked in any way: see Appendix). For example, the idiom *sweep something under the carpet* (or, using a frequent lexicographic placeholder convention, *sweep sth under the carpet*) was presented as *Old age is a subject that is swept under the carpet*.

PROCEDURE

The exercise was conducted in a classroom setting using pen and paper, under no specific time constraints. Participants were instructed (in Polish, their native language) that each sentence held 'a difficult expression' and that they should underline the word they would choose to look up in a dictionary. Participants worked individually and with no access to any other materials except the sheets with the test items (see Appendix). Participants were not rewarded or graded, nor did they sign their sheets. The sheets were then collected, and all underlines were transferred to a spreadsheet by a student assistant and double-checked by the first author. This formed the basis for data analysis.

DATA ANALYSIS

DATA CODING

For each of the component words in the MWE's tested which was underlined at least once, the four factors as in TABLE 1 were checked and coded, where applicable, with their corresponding letter(s) in the relevant column of the database. The codes were entered also in the case of tied items: for example, if there were two nouns in the MWE, one of which was the underlined item, the code *s* was entered for each of these nouns. Then, whether any such tied cases were counted as successfully predictive, depended on the specific type of analysis: *no ties* or *with ties* (see below). The coding was done by a properly qualified outside researcher based on instructions as in TABLE 1 and then independently by the two authors, with perfect inter-rater agreement.

TABLE 1. Coding scheme for predictive factors

f	frequency	entered if the word underlined was the **least frequent** in the MWE: (1) by frequency band as represented by the number of stars given in the relevant entry in the *Macmillan English Dictionary* (Rundell, 2007) (2) by the frequency indicated in the Corpus of Contemporary American English (COCA), lemmatized (Davies, 2008-)
s	POS (syntactic class)	entered if the POS (syntactic class) of the word underlined was **most preferred** out of all the component words of this MWE as per findings from Lew (2012); the order of preference is: N > Adj > V > Adv > all others (see Figure 1 in Lew 2012)
i	initial position	entered if the word underlined was the **first content word** in the MWE (content word = noun, verb, adjective, or (lexical) adverb)
l	length	entered if the word underlined was the **longest word by letter count** in the MWE

As many of the factor codes were entered in the database as there were valid factors that selected the particular word within this MWE; therefore, each component word of each MWE could get from zero to four-letter codes, in any combination. This may sound fairly confusing, so we illustrate the coding process in a worked example below.

A WORKED EXAMPLE

MWE: *a leopard can't change his spots*
Word underlined: *leopard*
This word is:

- the least frequent word in the MWE (f)
- the first content word in the MWE (i)
- longest by letter count (l)
- most preferred syntactic class (= noun) amongst the words in the MWE (s).

Therefore, the word leopard in this MWE scored positive on **all four** of the criterial factors and would have the four letters "fils" entered in the spreadsheet.

DETERMINING USER CHOICES

As per instructions, our participants indicated their willingness to look up a particular item by underlining it on their paper test sheet. A systematic analysis of the patterns of underlined items and predictive power of the factors considered was then undertaken, using three approaches as follows:

1. best predictor of success with underlines grouped by MWE item: this means that for each test MWE item we tallied the success of each of the four factors considered across all participants, and under two sub-scenarios:
    a. no ties (orthodox, one-keyword only);
    b. with ties (relaxed, one or two keywords allowed).

2. individual participant strategy, with underlines grouped by participant: this time we tallied the success of each of the four factors for each participant, across all MWE items.

The scenario underlying the no-ties approach (point 1.a. above) was that the dictionary will treat a given MWE in one place only—under what the dictionary editors consider the most opportune headword—and in doing so the editors will apply a pre-defined criterion, such as lexical frequency. The with-ties variant analysis (point 1.b. above) relaxes the single-keyword requirement, allowing not just one, but also two headwords in those cases when a given criterion does not yield a single winner, yet the most frequently underlined item is amongst the co-winners (for example, two nouns in the MWE, one of which is the most frequently underlined). Finally, analysis (2) looks at individual participants, rather than MWE's, in each case isolating the dominant strategy of the individual participant, expressed in terms of our design factors.

In addition, analyses (1.a.) and (1.b.) were done for all participants pooled, as well as separately within the two participant groups: that is the more advanced university students (U) and the less advanced high school students (HS). Analyses were conducted in the R environment for statistical computing (R Core Team, 2022), while initial data entry was done in Microsoft Excel sheets.

RESULTS

Participants' choices turned out to be quite varied. To give readers a sense of the range of variation within and across items, TABLE 2 lists the typical eight items with their choices. For example, in the first MWE listed in the table (*live happily ever after*), 33 participants underlined the word *live*,

58 participants underlined the word *happily*, 21 participants underlined *ever*, and 14 underlined the word *after*. As can be seen, there are rarely obvious 'winners' in terms of items attracting a very clear majority of choices, and in some cases (e.g., *live happily ever after*) user preferences are widely divergent.

TABLE 2. A sample of eight test items; numbers in parentheses indicate how many participants underlined the immediately preceding word

live (33) happily (58) ever (21) after (14)
rack (118) her (0) brains (30)
wash (28) dirty (8) linen (112) in (0) public (3)
a (0) leopard (72) can't (0) change (8) his (0) spots (69)
cut (22) off (0) your (0) nose (40) to (0) spite (83) your (0) face (5)
blow (50) the (0) whistle (100) on (0) you (0)
chickens (52) are (0) coming (3) home (1) to (0) roost (95)
a (0) nasty (68) piece (41) of (0) work (29)

ANALYSIS BY MWE: WHICH FACTOR BEST PREDICTS USER CHOICES

NO TIES: ONE KEYWORD ONLY

This part of the analysis will focus on the one keyword in the multi-word expression that attracted the most user underlines. This approach corresponds to the traditional lexicographic policy of giving full lexicographic treatment under one headword only. FIGURE 1 reports the number of MWE's, out of a total of 37, with correctly predicted most frequently underlined choice by each of the four design factors: frequency (band); (initial) position in MWE; length in letters; POS (syntactic class). Any ties are penalized in this analysis, that is tied cases do not count as correctly predicted. As seen in FIGURE 1, **frequency band** is the single best predictor of top user underlines, covering over two-thirds (68%) of MWE's when both user groups are considered jointly, labelled *All* in the Figures (two-sided Fisher's Exact Test, **frequency band** compared to **length** or **position**, Odds Ratio (*OR*) = 2.7, 95% Confidence Interval for the Odds Ratio ($CI_{95\%}$) = [1.0, 7.4], $p = 0.06$). The prediction success rate goes up to 73% for the HS group.

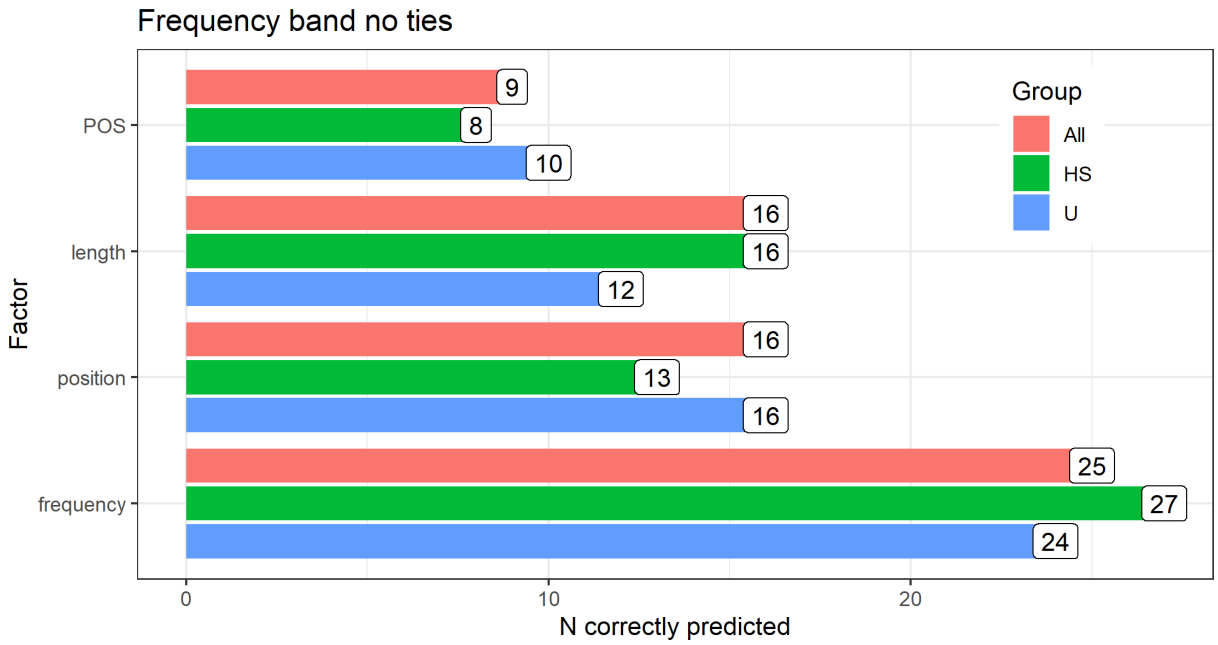

FIGURE 1. Number of MWE's (out of a total of 37) with correctly predicted top underlines by the four factors

SIX CASES OF TIES: BEST TIEBREAKERS

In the above analysis, there remain six MWE's which are tied, or unresolved: in those cases, the frequency band factor identifies more than one candidate keyword in the MWE because there is more than one component word in the MWE that happens to be in the same lowest frequency band. Therefore, if a lexicographer is restricted (by editorial policy) to just one keyword under which to list a given MWE, then the frequency band criterion alone will not provide a single solution.

To see if these ties can be broken using exact frequency ties, rather than categorical frequency bands, we extracted exact frequency counts from the *Corpus of Contemporary American English* (Davies, 2008-). We also considered the three remainings (non-frequency) criteria as potential tiebreakers. These results are presented in TABLE 3, which gives the count—out of a total of six—items that can be successfully resolved using a particular tiebreaker as a decision criterion.

TABLE 3. Number of tied MWE's which get correctly resolved using potential tiebreakers

Tiebreaker	***All***	***U***	***HS***
less frequent (COCA)	5	2	5
earlier	2	2	2
longer	4	1	4
POS	3	2	3

As seen in TABLE 3, using the more fine-grained COCA frequency counts—rather than the broader frequency bands from the *Macmillan English Dictionary* (MED)—we were able to resolve correctly (i.e., identify the top underlined words) five out of six problematic cases (two-sided Fisher's Exact Test, **exact frequency** compared to **length** or **position**, Odds Ratio (*OR*) =

5.5, 95% Confidence Interval for the Odds Ratio ($CI_{95\%}$) = [1.8, 17], p = 0.001). The single case in which the exact corpus frequency count does not lead to resolution was:

hair (74) of (0) the (0) **dog** (36)

—where the word with a lower COCA frequency count, *dog*, received only about half of the underlines of *hair*, a slightly more frequent word in COCA. As the numbers indicate, the high success of exact frequency counts in resolving the ties holds for the complete pool of participants, as well as for the HS group alone. However, this high level of success does not apply to university attendees: in this group, there remained three additional MWE's in which the exact frequency count did not resolve. The least frequent word—though not the one most frequently underlined by university students—is bolded:

live (27) **happily** (22) ever (14) after (13)
a (0) **greasy** (33) spoon (45)
stop (16) dead (36) in (0) her (0) **tracks** (26)

All in all, when the exact frequency is used rather than the frequency band, prediction accuracy improves, though mostly for HS (up to 89% of MWE's). The details are given in FIGURE 2, which gives the number of MWE's (out of a total of 37) with correctly predicted top-choice underlines by the four criterial factors: exact frequency from the Corpus of Contemporary American English (COCA); (initial) position in MWE; length in letters; syntactic class (POS). This analysis is different from the one in FIGURE 1 in that exact frequency counts are used rather than frequency bands.

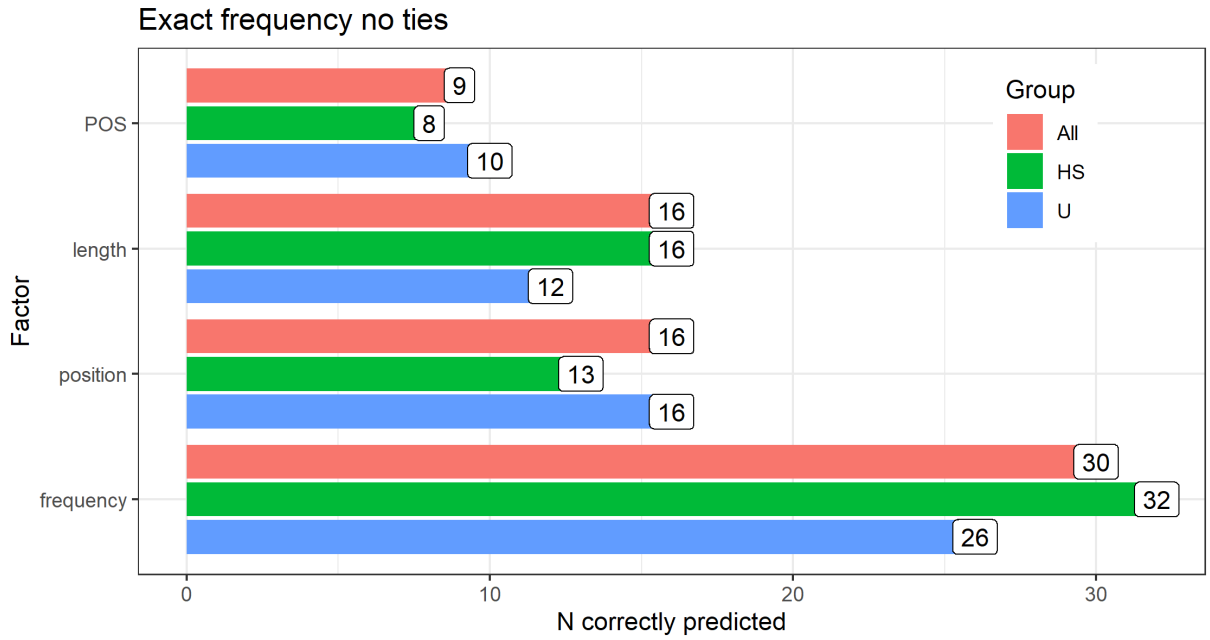

FIGURE 2. Number of MWE's with correctly predicted top underlines. Same as in FIGURE 1 except exact frequency is used

RELAXING THE SINGLE KEYWORD RESTRICTION: TIES ALLOWED

Thus far, our analysis has assumed that a single best keyword must be selected, in line with an orthodox editorial policy of including a given MWE under a single entry. Consequently, when a given criterion such as the lowest frequency band pointed to more than one component word in the MWE, it was deemed unable to identify the winner even if the most frequently underlined word *was* in the lowest frequency band. In such cases, the factor was penalized for this MWE (i.e., classified as a failure). However, users' choices can be quite variable: consider cases like the ones given in TABLE 4.

TABLE 4. Two examples of MWE items with no clear single preferred keyword

U	HS
blind (27) **drunk** (51)	**blind** (42) drunk (24)
live (27) happily (22) ever (14) after (13)	live (6) **happily** (36) ever (7) after (1)

Given this variability, it makes sense to explore what happens when the orthodox one-keyword restriction is relaxed by allowing ties of up to two keywords. For example, this means that if the frequency band criterion identifies two component words in the lowest frequency range, and one of the words happens to have received the most underlines, then this would now be counted as a point in favour of the criterial factor. In terms of lexicographic practice, this corresponds to allowing a single MWE to be listed under *two* headwords—a policy not altogether uncommon in English pedagogical lexicography (albeit not necessarily conscious or coordinated, as evidenced by the occasional different treatment of the same idiom in two different places). Apart from relaxing the no-ties restriction, everything else remains the same, and the corresponding numbers are now given in FIGURE 3. The FIGURE thus gives the number of MWE's (out of a total of 37) with correctly predicted top-choice underlines by the four criterial factors: frequency (band); (initial) position in MWE; length in letters; syntactic class (POS). This time, however, ties up to two are allowed.

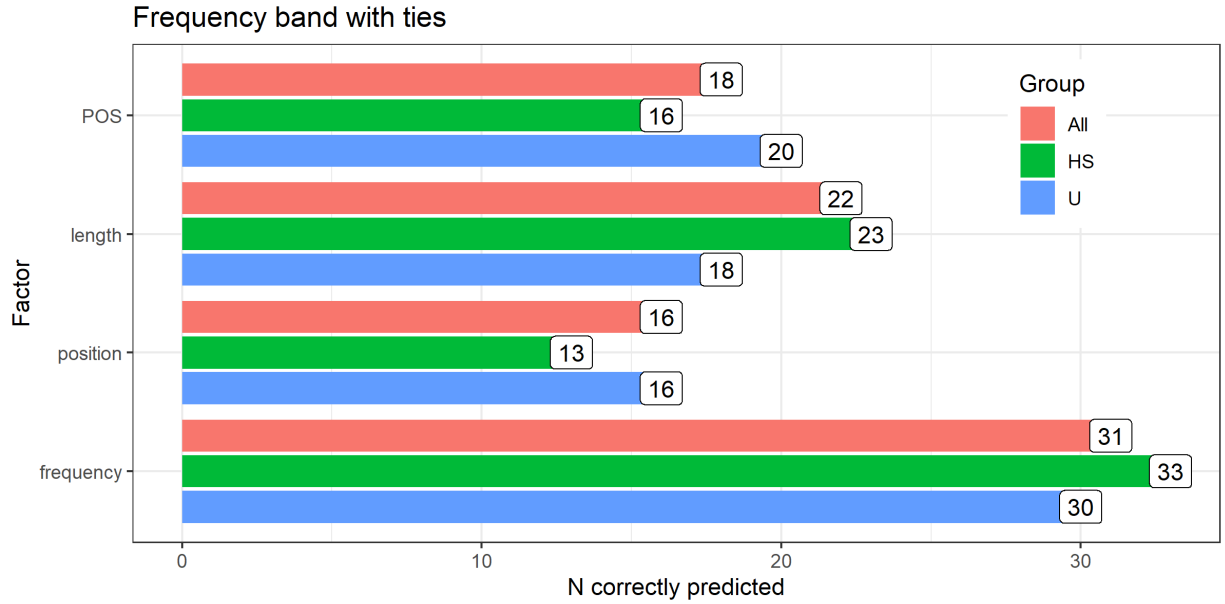

FIGURE 3. Number of MWE's with correctly predicted top underlines; ties up to two are allowed

Comparing FIGURE 1 and FIGURE 3, it will be seen that the success rates have now improved, reaching 84% overall, and as high as 89% in the HS group, for the frequency band factor. In this scenario, adding COCA exact frequency does not lead to improvement and will not be presented here. (In fact, COCA performs marginally worse, presumably because the frequency counts used in the *Macmillan English Dictionary* are a better match for the language experience of our learners of English).

PARTICIPANT STRATEGIES

Another way to examine the participants' look-up preferences is to treat each participant separately and compute that participant's dominant consultation strategy, across the 37 items, expressed in terms of our design factors. These results are presented in FIGURE 4. In both the U and HS groups, the most popular strategy (represented by about 80% of all participants) appears to be linked to the frequency criterion, with students preferring rare words as top candidates for look-up. A minority of nearly 20% (slightly less amongst the HS students) seem to be guided by position, focusing on the initial content word of the problematic expression (two-sided Fisher's Exact Test, **frequency** compared to **position**, Odds Ratio (*OR*) = 4.3, 95% Confidence Interval for the Odds Ratio ($CI_{95\%}$) = [2.7, 7.0], $p < 0.001$).

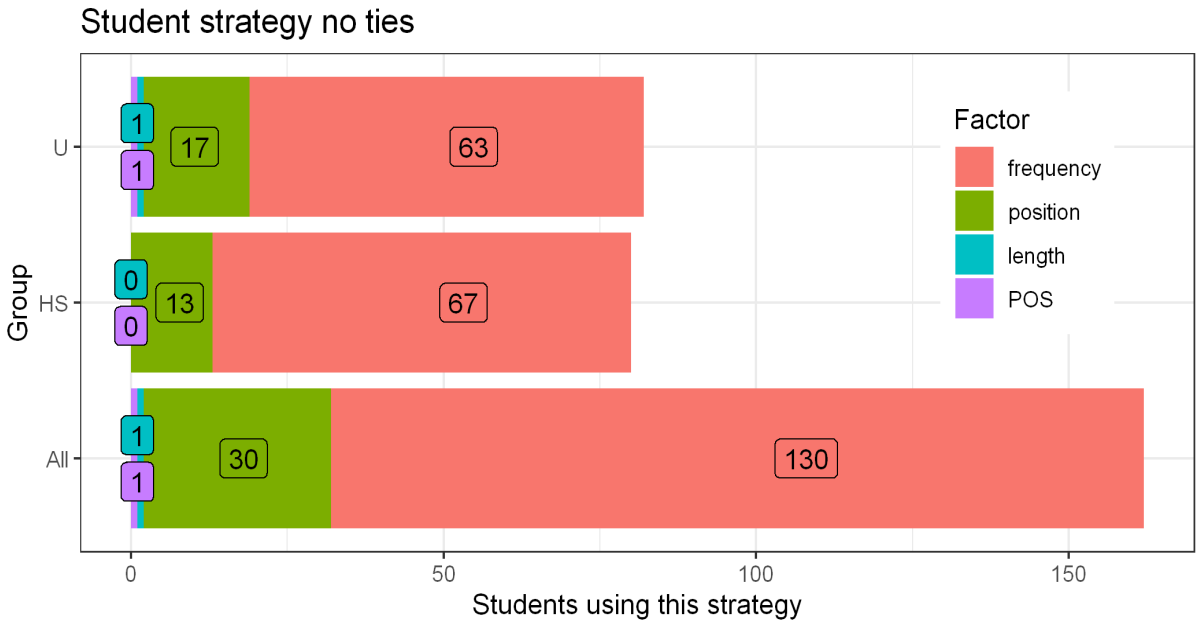

FIGURE 4. Dominant strategies of student participants. For most participants, choices are best explained using the frequency band

DISCUSSION

The present study has confirmed lexical frequency (Lew, 2012) as the leading factor explaining and predicting the choices made by student participants when challenged with semantically difficult multi-word expressions. While the look-up behaviour remains in some parts haphazard and unpredictable, it does seem possible to predict the typical choices with useful accuracy. For

our medium-sized sample of 37 multi-word expressions, 25 (or over two-thirds) would be consulted via the single least frequent component word, using the MED frequency bands. Using COCA exact corpus frequencies whenever there were two words in the same least frequent band boosted the accuracy to 30 items, or 81%. However, when the orthodox restriction to one keyword only is relaxed, the accuracy goes up to 31 items (84%). This means users would be able to find the relevant expression under the first entry they consult 84 per cent of the time, provided the dictionary includes treatment in two places for those cases where the frequency criterion does not point to a single least frequent keyword. In the no-ties approach, an improvement from using exact frequency counts is particularly noticeable for the lower-level HS students, but not so much for university-level students. This is interesting in itself. One possible explanation might be that HS students, with less vocabulary exposure, might be more sensitive to smaller frequency differences between what are all *familiar* words to university students, though for now this remains speculation.

All in all, allowing the inclusion of an MWE under two headwords when there are two equally rare component words seems like a good lexicographic policy that we would currently recommend (but see the section on digital dictionaries below). Overall, corpus frequency is a good indicator of user behaviour, presumably because rare words are likely to be unfamiliar to dictionary users, and those are the ones they would want to try following up in reference works. Naturally, individual language experiences must play a role, with each person's perception of frequency being unique up to a point, yet frequency counts derived from appropriate corpora that are used by leading dictionaries for learners still do a good job.

Other criterial factors considered in this study turn out to have lower predictive value than frequency. This is the first study to factor in word length: greater length and initial position vie for second place, with the length being more useful than position under the "ties allowed" scenario. Part of speech turns out to be the least predictive of the four factors considered. This study is also the first one to provide evidence on MWE's embedded in actual sentences, rather than bare expressions presented in isolation. However, the findings largely coincide with those from previous studies.

USER CHOICES VERSUS WHAT DICTIONARIES DO

Since we collected empirical data on 37 expressions, we thought it might be revealing to see to what extent leading dictionaries for English learners match actual user choices. To do that, we looked up all our MWE's in recent print copies of the five leading dictionaries for learners of English (the *Big Five*). These results are given in FIGURE 5, indicating for each dictionary the presence and type of lexicographic treatment under the most frequently underlined component word. *Full treatment* refers to the complete lexicographic treatment of the MWE under the component word most preferred by our users. *Cross-reference* means restricted treatment that directs the dictionary user to another entry holding full treatment. *Not listed* means our target expression is completely absent from the entry, as the editors chose to treat the MWE under another component word (we should stress that all of our target expressions were included in all five dictionaries).

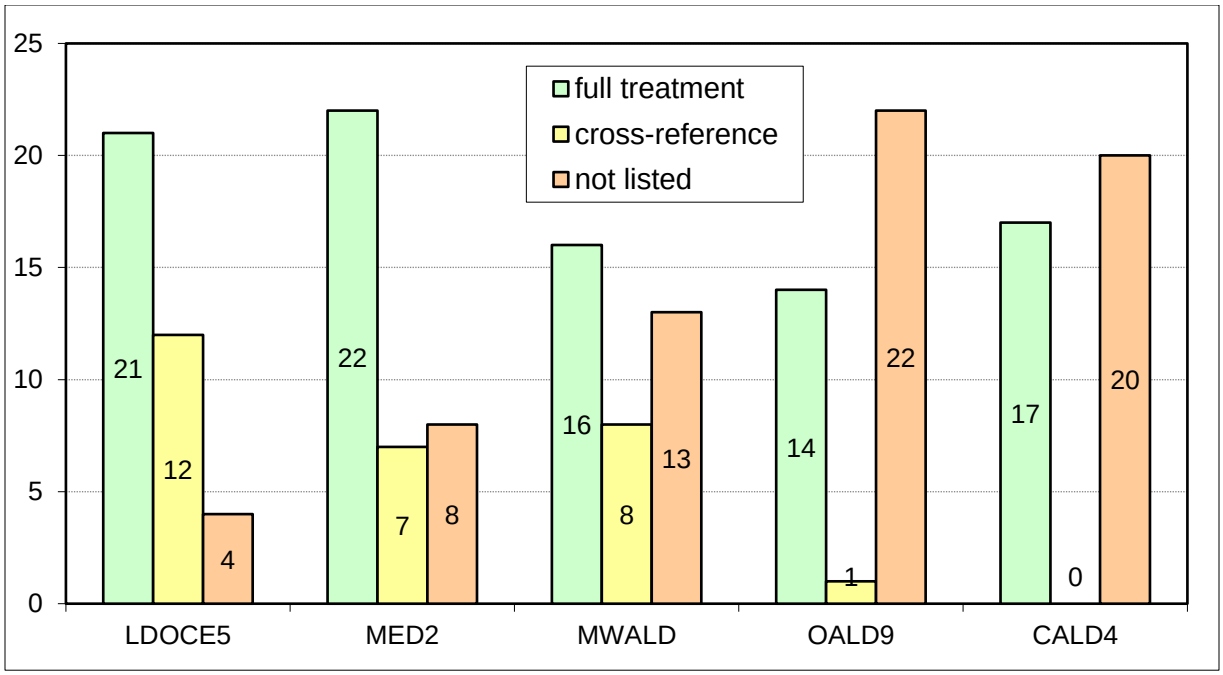

FIGURE 5. Accuracy of the leading five dictionaries for learners of English in listing target MWE's under the most preferred component word

As seen in FIGURE 5, the *Longman Dictionary of Contemporary English* (Marvick, 2014) and *Macmillan English Dictionary* (Rundell, 2007) are most closely aligned with our users' preferences. *Oxford Advanced Learner's Dictionary of Current English* (Deuter & Bradbery, 2014) and *Cambridge Advanced Learner's Dictionary* (McIntosh, 2013) do much worse in this respect, with *Merriam-Webster's Advanced Learner's English Dictionary* (Perrault, 2008) in the middle. What is worthy of note is that even the best-performing *Macmillan English Dictionary* (full treatment under the best component word for 22 expressions out of the 37) would have done much better, had it adopted the simple criterion of lexical frequency (compare the relevant numbers in FIGURE 1, FIGURE 2, and FIGURE 3).

CONCLUSION

Admittedly, the central issue under discussion—the choice of entry under which to place a multi-word expression—is most directly relevant in print dictionaries. In such dictionaries, space is restricted, and any duplication of information is avoided as far as possible. If a restrictive policy is adopted of treating a multi-word expression under a single entry only, it is crucially important that the entry is one which most users would choose to consult. However, our recommendation arising from this study is to allow treatment under two headwords in multi-word expressions that include two infrequent component words. There is also the option of restricted treatment in the form of cross-references, which could be used to some degree.

In digital dictionaries, it is technically possible to store a multi-word expression in a single place while at the same time displaying it in more than one entry. However, this is not necessarily ideal, and certainly not for entries that, by their very nature, are already quite substantial. For

example, it is quite clear that it would not be wise to include *hair of the dog* under the function words *of* or *the*. Storing information is one thing; presenting it legibly is another (Lew, 2010): presentation space is limited also in digital dictionaries, especially those equipped with smaller screens, such as smartphone dictionary apps.

The available options are still essentially (1) full treatment, (2) restricted treatment (cross-reference), and (3) no treatment. One of the strengths of digital dictionaries vis-à-vis print dictionaries is that cross-referencing can be implemented as a hyperlink, at a much smaller cost to users than an old-fashioned paper entry search. Therefore, it makes sense to adopt cross-referencing as the option of choice. Further, although entry readability and data overload (or clutter, if you like) are still a concern, the cost of relaxing the orthodox single-headword policy is less. For digital dictionaries, then, the 'with ties' options are most directly applicable. All in all—despite what might appear at first sight—the issue of MWE placement addressed in the present contribution remains relevant also for digital dictionaries.

ACKNOWLEDGEMENTS

The Authors would like to thank Małgorzata Mazańska for help with data collection.

REFERENCES

Atkins, B. T. S., & Knowles, F. E. (1990). Interim report on the EURALEX/AILA Research Project into Dictionary Use. In T. Magay & J. Zigány (Eds.), *Budalex '88 Proceedings* (pp. 381–392). Akadémiai Kiadó.

Atkins, B. T. S., & Rundell, M. (2008). *The Oxford guide to practical lexicography*. Oxford University Press.

Atkins, B. T. S., & Varantola, K. (1998). Language learners using dictionaries: The final report on the EURALEX/AILA Research Project on Dictionary Use. In B. T. S. Atkins (Ed.), *Using dictionaries. Studies of dictionary use by language learners and translators* (pp. 21–81). Niemeyer.

Béjoint, H. (1981). The foreign student's use of monolingual English dictionaries: A study of language needs and reference skills. *Applied Linguistics. 2*(3), 207–222. C:\3\LIBRARY\Lexicography\Bejoint_1981.htm

Bogaards, P. (1990). Oú cherche-t-on dans le dictionnaire? *International Journal of Lexicography. 3*(2), 79–102. http://ijl.oxfordjournals.org/cgi/reprint/3/2/79

Bogaards, P. (1991). Word frequency in the search strategies of French dictionary users. *Lexicographica. 7*, 202–212. http://dx.doi.org/10.1515/9783110244113.202

Bogaards, P. (1992). French dictionary users and word frequency. In K. H. Tommola, K. Varantola, T. Salami-Tononen & J. Schopp (Eds.), *EURALEX '92 Proceedings* (pp. 51–59). Department of Translation Studies, University of Tampere.

Chen, P., & Zhao, C. (2022). The treatment of academic lexical bundles in online English monolingual learners' dictionaries. *International Journal of Lexicography*. https://doi.org/10.1093/ijl/ecab032

Dalpanagioti, T. (2021). *A cognitive perspective on the representation of MWEs in electronic learner's dictionaries.* Paper presented at the Electronic lexicography in the 21st century: post-editing lexicography. Proceedings of the eLex 201 conference.

Davies, M. (2008-). The Corpus of Contemporary American English. http://corpus.byu.edu/coca/

Deuter, M., & Bradbery, J. (2014). *Oxford Advanced Learner's Dictionary of Current English* (Ninth ed.). Oxford University Press.

Farina, D. M. T. C., Vrbinc, M., & Vrbinc, A. (2019). Problems in online dictionary use for advanced Slovenian learners of English. *International Journal of Lexicography. 32*(4), 458-479. https://doi.org/10.1093/ijl/ecz017

Granger, S., & Paquot, M. (2008). Disentangling the phraseological web. In S. Granger & F. Meunier (Eds.), *Phraseology: An Interdisciplinary Perspective*. John Benjamins.

Heuberger, R. (2020). Monolingual online dictionaries for learners of English and the opportunities of the electronic medium: A critical survey. *International Journal of Lexicography. 33*(4), 404-416. https://doi.org/10.1093/ijl/ecaa018

Knežević, L., Halupka-Rešetar, S., Miškeljin, I., & Milić, M. (2021). Millennials as dictionary users: A study of dictionary use habits of Serbian EFL students. *SAGE Open. 11*(2), 21582440211008422. https://journals.sagepub.com/doi/abs/10.1177/21582440211008422

Lew, R. (2010). Multimodal lexicography: The representation of meaning in electronic dictionaries. *Lexikos. 20*, 290–306. http://lexikos.journals.ac.za/pub/article/download/144/85

Lew, R. (2012). The role of syntactic class, frequency, and word order in looking up English multi-word expressions. *Lexikos. 22*, 243–260. http://lexikos.journals.ac.za/pub/article/view/1006/523

Marvick, K. C. (2014). Longman Dictionary of Contemporary English *LDOCE6* (6th ed.). Longman.

McIntosh, C. (2013). Cambridge Advanced Learners' Dictionary *CALD4* (4th ed.). Cambridge University Press.

Parizoska, J., & Petrovic, I. F. (2020). Idiom variations in online dictionaries. *Rasprave. 46*(2), 941-955

Perrault, S. J. (2008). Merriam-Webster's Advanced Learner's English Dictionary *MWALED*. Merriam-Webster.

R Core Team. (2022). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org

Rundell, M. (Ed.) (2007) MED2 (2nd ed.). Macmillan.

Tomczak, E., & Lew, R. (2019). "The Song of Words": Teaching Multi-Word Units with songs. *3L: Language, Linguistics, Literature. 25*(4), 16-33. http://doi.org/10.17576/3L-2019-2504-02

Tono, Y. (1987). *Which word do you look up first? A study of dictionary reference skills.* (M.Ed. Thesis), Tokyo Gakugei University, Tokyo.

Wojciechowska, S. (2020). Access routes to BODY PART multiword expressions in the 'Big Five' MELDs: Use of hyperlinks. *Lexikos. 30*(1)

Zipf, G. (1949). *Human Behavior and the Principle of Least Effort*. Addison-Wesley.