

Academic Word Families in Online English Dictionaries

Geraint Paul Rees, *Translation and Language Sciences,*
Pompeu Fabra University, Barcelona, Spain
(geraintpaul.rees@upf.edu) (<https://orcid.org/0000-0002-9204-8073>)

Abstract: The concept of the word family has been widely employed in research on vocabulary in the teaching and learning of foreign and second languages. The underlying assumption being that once learners know one member of a word family, they can recognise other members. Empirical research supports this vis-à-vis receptive knowledge of inflectionally related wordforms. However, studies of academic writing indicate that using appropriate derivative forms of a known word is challenging, suggesting a need for dictionaries with morphological support for writers. Traditionally, in paper-based dictionaries, this need could not be fulfilled due, in part, to space constraints. This study aims to establish if it is met in five online English dictionary websites. It analyses the treatment of seventy-four academic wordforms which academic writers have been shown to have difficulty deriving when presented with the related base word. Results indicate good coverage of the derivative forms across the dictionary websites examined but inconsistency within and between resources in the way in which forms are treated. Differences include the status as entries or subentries and the provision of writing support features such as examples, grammar patterns, and collocation information. Finally, changes to the treatment of derivatives to better serve academic writers are suggested.

Keywords: ACADEMIC WRITING, DERIVATIVE FORMS, LEXICOGRAPHY, MORPHOLOGY, ONLINE DICTIONARIES, VOCABULARY ACQUISITION, WORD FAMILIES, WRITING SUPPORT

Opsomming: Akademiese woordfamilies in aanlyn Engelse woordeboeke.

Die woordfamilie-konsep is reeds wyd in woordeskatnavorsing in die onderrig en aanleer van vreemde en tweede tale ingespan. Die onderliggende aanname word gemaak dat wanneer leerders een lid van 'n woordfamilie ken, hulle ook ander lede kan herken. Empiriese navorsing steun hierdie aanname ten opsigte van reseptiewe kennis van fleksieverwante woordvorme. Studies van akademiese skryfwerk toon egter dat die gebruik van toepaslik afgeleide vorme van 'n bekende woord 'n uitdaging bied, wat daarop dui dat daar 'n behoefte aan woordeboeke met morfologiese steun vir skrywers bestaan. Tradisioneel kon, deels weens ruimtebeperkings, nie aan hierdie behoefte in papiergebaseerde woordeboeke voldoen word nie. In hierdie studie word beoog om vas te stel of daar in vyf aanlyn Engelse woordeboekwebtuistes wel hieraan voldoen word. Die hantering van vier-en-sewentig akademiese woordvorme waarmee akademiese skrywers sukkel om afleidings daarvan te vorm wanneer hulle die verwante basiswoord teëkom, word geanaliseer. Die resultate dui op goeie verteenwoordiging van die afgeleide vorme in die woordeboekwebtuistes wat ondersoek is, maar toon ook teenstrydighede binne en tussen hulpbronne t.o.v. die metode waarop die vorme hanteer word. Verskille sluit die status as inskrywings of subinskrywings en die voorsiening

van skryfhulpkenmerke soos voorbeelde, grammatikale patrone en kollokasie-inligting in. Ten slotte word veranderings aan die hantering van afleidings voorgestel om akademiese skrywers beter van hulp te kan wees.

Sleutelwoorde: AKADEMIESE SKRYFWERK, AFGELEIDE VORME, LEKSIKOGRAFIE, MORFOLOGIE, AANLYN WOORDEBOEKE, WOORDESKATVERWERWING, WOORDFAMILIES, SKRYFHULP

Introduction

Over the last three decades, the term 'word family' has been used in language teaching and vocabulary research to describe the categorisation of wordforms based on their inflectional and derivational morphology. The construct has been adopted enthusiastically in research on vocabulary in English language teaching. A key factor motivating the concept of word family (henceforth WF) was a desire to provide guidelines for the treatment of morphologically related wordforms in lexicography and language teaching (Bauer and Nation 1993). The starting point for this study is a list of seventy-four wordforms frequently used in academic contexts. It comprises sixteen basic wordforms and their derivatives. Empirical research has shown that L2 users have difficulty producing the WF members (i.e., related wordforms) of these forms in writing (Schmitt and Zimmerman 2002). This study aims to establish how well L2 users of the "Big Five" English dictionary websites are supported when producing these problematic forms by examining their treatment on these websites.

Word families and levels. The idea motivating WFs is that wordforms can be grouped based on their inflectional and derivational morphology. These groups can then be organised into levels. Table 1 reproduced from Bauer and Nation (1993: 254) shows the levels for the WFs *develop*, *wood* and *bright*.

Table 1: Additions to a WF at different levels of inflection and affixation (reproduced from Bauer and Nation (1993: 254))

Word families			
2	develop develops developed developing	wood wood's woods wooded	bright brighter brightest
3	developable undevelopable developer(s) undeveloped	woody woodiest woodier woodiness	brightly brightish brightness

4	development(s) developmental developmentally		
5	developmentwise semideveloped antidevelopment	wooden	brighten
6	redevelopment predevelopment	anti-wood	

An increase in WF level entails greater formal or semantic irregularity. At Level 1, each form represents a distinct word (i.e., one word = one family). At Level 2, inflected forms with the same base are grouped. The idea being that a learner who can recognise and use *develop* or any of its inflected forms could recognise and use the base or any of its other inflected forms. From Levels 3 to 6 eight criteria determine the level of an affix and its derived wordform (Bauer and Nation 1993: 256).

1. Frequency (generalisability): Affixes at lower levels occur in many wordforms. For example, the Level 2 inflectional affixes *-s*, *-ed*, *-ing*, are common to all English verbs. In contrast the affixes *pre-* and *re-* are far less generalised.
2. Productivity: The possibility of the affix forming new wordforms. Inflectional affixes *-s*, *-ed*, *-ing* frequently produce new forms with the base of any verb. Whereas since *-ful* is far more selective in the nouns and verbs it combines with, it produces far fewer wordforms.
3. Predictability: The extent to which the meaning of the word created by affixation can be predicted from the meaning of the base and the affix. For example, *-ly* attached to an adjective X, typically means 'in X manner', thus is highly predictable. In contrast, *-ful* when attached to nouns does not always produce word with predictable meaning (e.g., *awful weather* ≠ *awe inspiring weather*).
4. Regularity of the written form of the base: At lower levels, removing the affix leaves the base intact, at higher levels orthographic changes to the base are evident (c.f. *red +ness* and *impose + ition*).
5. Regularity of the spoken form of the base: At lower levels removing the affix leaves it phonologically intact, at higher levels phonological accommodations are evident. For example, removal of the Level 6 affix *-ify* from *mystify* gives *myst*; not a free base in its spoken form.
6. Regularity of spelling of the affix (allomorphy): For example, *pre-* has one written form, while *in-*, *im-*, *il-*, and *ir-* are allomorphs of *in-*.
7. Regularity of the spoken form of the affix (allomorphy): The extent to which the phonological form of the affix is predictable. For example, although the Level 1 affix *-ed* has three spoken forms, these are predictable.

8. Regularity of function: The extent to which the affix attaches to a base of a particular word class and produces a word of a particular class. For example, *-ship* always combines with nouns to produce nouns.

By applying the criteria above, Bauer and Nation (1993) produced the list of affixes in Table 2. Two levels are omitted: Level 1 where each wordform is treated as a different WF, and Level 7 where items have classical roots and affixes that are not found in the sample of WFs in this study.

Table 2: Affixes at different WF levels

Level	Affixes
2	<i>-s, -ed, -ing</i>
3	<i>-able, -er, -ish, -less, -ly, -ness, -th, -y, non-, un-</i>
4	<i>-al, -ation, -ess, -ful, -ism, -ist, -ity, -ize, -merit, -ous, in-</i>
5	<i>-age, -al, -an, -ance, -ant, -ary, -atory, -dom, -eer, -en, -en, -ence, -ent, -ery, -ese, -esque, -ette, -hood, -i, -ian, -ite, -let, -ling, -ly, -most, -ory, -ship, -ward, -ways, -wise, ante-, anti-, arch-, bi-, circum-, counter-, en-, ex-, fore-, hyper-, inter-, mid-, mis-, neo-, post-, pro-, semi-, sub-, un-</i>
6	<i>-able, -ee, -ic, -ify, -ion, -ist, -ition, -ive, -th, -y, pre-, re-</i>

Word families and language teaching and learning. The usefulness of WFs for language teaching relies on the assumption that once learners know one member, they can recognise others. This has been termed relational knowledge (Tyler and Nagy 1989). Some empirical research supports this for L1 readers and inflectionally related wordforms. However, that derived forms are generally acquired after inflected forms suggests they pose greater problems (Berko 1958). For L2 users, the assumption of relational knowledge is more uncertain. Even proficient L2 users find using suitable derived forms of a known word challenging.

Studies on L2 writing or vocabulary acquisition suggest that learners find derivational morphology challenging. A longitudinal study of English vocabulary acquisition involving three L2 English postgraduate students in the UK indicated gaps in participants' morphological repertoire, particularly regarding the formation of adjectives and adverbs. Schmitt (1998) suggests that morphological errors become fossilised since two of the three participants made little progress producing morphologically related forms over an academic year. Another study of the English word association and grammatical suffix knowledge of 95 secondary and undergraduate students of English in Japan found participants gained 330 words on average over an academic year but could only produce 15% of the possible derivatives (Schmitt and Meara 1997). Similarly, in a study

of TOEFL vocabulary involving 30 learners taking English language courses in preparation for undergraduate study in the UK, participants could only produce derivatives in all four major word classes for 12 of 180 possible target words (Schmitt 1999).

Research focusing on productive knowledge of derivational morphology among learners is rarer. Schmitt and Zimmerman's (2002) carefully designed study examined the productive knowledge of 106 L2 English students who comprised two groups: One undertaking pre-sessional and undergraduate English language courses at universities in the US and the UK, and another an MA in English language teaching at a university in the UK. Participants were given 16 prompt words for which they were asked to complete gapped sentences by producing derivative forms of the prompt word from the four major word classes (noun, verb, adjective, and adverb). Participants produced only 50% of the derivative forms permissible. Although the presumably more proficient MA group performed better, knowledge of derived word-forms was still partial even for words which participants felt they knew well. This demonstrates a need for dictionaries that support written production of derived wordforms.

Word families and dictionary making. WFs were posited to help lexicographers treat morphology in a principled and consistent way. Bauer and Nation (1993) criticise the inconsistent treatment of derived forms as entries and sub-entries in several general-purpose English dictionaries from the late 1970s and 1980s. They are not alone in highlighting this issue. However, much research has focused on affixes themselves rather than the derivative forms produced by affixation. For example, Stein (1985) highlights different policies on the positioning of affixes in the indexes of several MLDs. Considering dictionaries as writing aids, it makes little sense to focus on affixes themselves rather than derivative wordforms produced by affixation. Writers are unlikely to ask, 'What word can I form with *-ize*?' but will likely query the use of a particular word, for example, 'How do I use *philosophize* in a sentence?'

There is some consensus on the treatment of wordforms derived by affixation. To be included, a derivative form must be established enough to occur above a certain frequency (De Caluwe and Taldeman 2003; Stein 1985). Semantic predictability is another important consideration: "The more the meaning of a combination is assumed to be inferable from the meaning of its constituents listed in the dictionary and the process of formation itself, the stronger the likelihood that it will not be listed as a dictionary item" (Stein 1985: 38).

Analyses of entries for derivative forms reveal diverse interpretations of these criteria. In an examination of eight monolingual English desk dictionaries including MLDs, Stein (1985) highlights inconsistent definition of *-ish* derivatives from adjectives designating colour, and inconsistent treatment of derived forms as either main entries or run-ons. Similarly, De Caluwe and Taldeman (2003) demonstrate inconsistent treatment of wordforms derived from *water* in the *Woordenboek der Nederlandsche Taal* (Van Sterkenburg 1992: 115), noting that some

are listed as separate entries or lemmas and others within the headword *water*.

WFs were posited to remedy these inconsistencies. The idea is that as formal and semantic irregularities increase with higher-level word families, they require "more attention" from the lexicographer (Bauer and Nation 1993: 255). Bauer and Nation suggest ignoring regular, semantically transparent word-forms at Level 1; listing those created by inflection affixation at Levels 2 and 3 as non-defined sub-entries and treating higher-level items as main entries.

WFs and electronic lexicography. Electronic lexicography has been suggested as the solution to the inconsistent treatment of derived forms. Firstly, ostensibly freed from space constraints of paper dictionaries, electronic dictionaries have the potential to include information on all the derived wordforms in a language¹. Secondly, unbound by the alphabetical index, they could offer several routes to the derivative wordform (De Caluwe and Taeldeman 2003). Regarding the first point, De Caluwe and Taeldeman (2003) stress the importance of not overwhelming the reader with information: "it is not the intention to confront the reader with an interminable amount of information, but this should be possible if the reader so desires" (De Caluwe and Taeldeman 2003: 121). Regarding access structure, they sketch an example of how an onomasiological query for "the fact/quality of being long" (De Caluwe and Taeldeman 2003: 123) might proceed in an ideal dictionary. With reference to *Elektronisches Lernerwörterbuch Deutsch-Italienisch/Dizionario Elettronico per Apprendenti Italiano-Tedesco* (ELDIT), Ten Hacken, Abel and Knapp (2006) present a detailed example of how derivative forms can be treated in electronic dictionaries.

Aims. Lexicography has changed significantly since Bauer and Nation's guidelines were published. Many space and alphabetical ordering constraints of paper dictionaries have been mitigated in online resources. These could feasibly accommodate calls from research on WFs in language teaching for greater writing support for L2 English with derived forms. Accordingly, this study aims to investigate how derivatives are represented in online English dictionary websites consulted by learners. It will answer the following research questions:

1. How well are derivationally related members of WFs covered by dictionary websites with online monolingual English dictionaries?
2. To what extent are they treated in a way which facilitates use in writing?

Methodology

In this section, the dictionary websites examined are discussed along with the reasons for their selection. Next, the sample of 74 derived wordforms shown to be problematic for L2 English users is presented and the process Schmitt and Zimmerman (2002) used to obtain this list explained. Finally, the categories and procedure used in this analysis are given.

Dictionary websites examined. This study examines the treatment of morphological behaviour on five popular English dictionary websites (Cambridge, <https://dictionary.cambridge.org/> (CAM); Collins, <https://www.collinsdictionary.com/> (CD); Longman, <https://www.ldoceonline.com/> (LONG); Macmillan, <https://www.macmillandictionary.com/> (MELD); and Oxford, <https://www.oxfordlearnersdictionaries.com/> (OX)²). The versions examined were those live in December 2022.

Monolingual Learners Dictionaries (MLDs) are the obvious place to investigate morphological information for learners. However, the migration from paper-based dictionaries to online dictionaries complicates this assumption. Of the "Big Five" monolingual English dictionary makers, only Longman and Macmillan offer direct access to their MLDs. *LDOCEonline.com* also gives access to the *Longman Business Dictionary*³ (LBD). Access to the MLDs of Cambridge, Collins, and Oxford is offered via portals which aggregate content from several different dictionaries. For example, the *collinsdictionary.com* entry for **precision** collates data from *Collins COBUILD* (COBUILD), *Collins English Dictionary* (CED)⁴, *Webster's New World College Dictionary* (Agnes 2010) (WNWCD⁴), and other ancillary sources. This study investigates the data presented by each portal rather than focusing only on entries from MLDs since, while dictionary researchers are cognisant of different dictionary types and their target users, many end-users, particularly those at lower proficiency levels, simply want to get the job done. It would be strange if an end-user disregarded information from *collinsdictionary.com* because it came from CED not COBUILD.

Productively challenging academic word families. Schmitt and Zimmerman (2002) judge 74 wordforms acceptable responses to gapped sentences based on sixteen prompt words. These represent an ideal sample with which to investigate the treatment of morphological information in online English dictionaries. The sixteen prompt words were selected from Coxhead's (2000) *A New Academic Wordlist* (AWL). This lends content validity since many English dictionary users, including those shown to have problems with derivative forms in the research discussed above, work in academic contexts.

To obtain the list of 74 acceptable derivate wordform responses, Schmitt and Zimmerman (2002) first extracted all listed derivatives from four learners' dictionaries⁵. Secondly, they used frequency information from the BNC1994 to remove infrequent derivatives. Finally, they elicited responses from 36 L1-English university students to the same gapped sentence prompts used by the non-native speakers. In arriving at their list of acceptable responses, they prioritised this final step. Table 3 shows WFs containing the basic and related wordforms along with their word class and WF level in parenthesis.

Table 3: List of problematic academic WFs (adapted from Schmitt and Zimmerman 2002: 168)

Noun	Verb	Adjective	Adverb
assumption (4)	<i>assume</i> (2)	assumed (2) X	X
<i>authority</i> (4) authorization (4)	authorize (4)	authorized (2) authoritive (6) authoritative (6)	authoritively (6) authoritatively (3)
tradition (2)	traditionize (4)	<i>traditional</i> (4)	traditionally (3)
selection (6)	<i>select</i> (2)	selective (6) select (2) selected (2)	selectively (3)
<i>access</i> (2)	<i>access</i> (2)	accessible (7) accessed (2)	accessibly (3) X
ethnicity (4)	X	<i>ethnic</i> (2)	ethnically (3)
philosophy (2)	philosophize (4)	philosophical (4) philosophic (6)	philosophically (3)
inevitability (4)	X	inevitable (3)	<i>inevitably</i> (3)
liberality (3) liberalization (4) liberalness (3) liberty (2)	liberalize (4)	<i>liberal</i> (4)	liberally (3)
<i>release</i> (2)	<i>release</i> (2)	released (2)	X
survival (4)	<i>survive</i> (2)	surviving (2)	X
<i>ideology</i> (2)	X	ideological (4) X	ideologically (3)
precision (6) preciseness (3)	X	<i>precise</i> (2)	precisely (3)
<i>minimum</i> (2) minimization (4)	minimize (4)	minimal (4) minimum (2)	minimally (3)
coherence (5) coherency (3)	cohere (2) X	<i>coherent</i> (2)	coherently (3)
persistence (5) persistency (3)	<i>persist</i> (2)	persistent (5)	persistently (3)

Note. Italics indicate the most frequently occurring member of each WF in BNC1994; X indicates Schmitt and Zimmerman's judgment that no typical form exists.

There are, at least, two notable points about this list. Firstly, Schmitt and Zimmerman (2002) treat *accessed*, *assumed*, *authorized*, *released* and *surviving* as adjectives. However, the first four could reasonably be verbs and *surviving* could be a verb or a noun. This is a frequent dilemma in English lexical analysis with no satisfactory answer (Hanks 2013). There are cases where these items are used as verbs and others where they are used as adjectives (Frankenberg-Garcia, Rees and Lew 2021). The analysis procedure below accounts for this. Secondly, the AWL has received criticism for ignoring discipline-specific differences in meaning, not accounting for the role of collocates in conditioning meaning and being based on the outdated *A General Service List* (West 1953) (Hyland and Tse 2007; Rees 2021). However, Schmitt and Zimmerman (2002) suggest these are words learners in academic contexts often need to produce. Their standout finding of partial knowledge of derivative forms demonstrates that L2 English users struggle to produce these words. Consequently, these are words for which they could conceivably consult a dictionary for guidance.

Procedure. A search for each of the seventy-four problematic derived forms is conducted on the five dictionary websites. The analysis of the results proceeds in two stages: Stage 1 records whether a wordform is covered; Stage 2 records whether the wordform is treated in a way that supports writing. Namely, whether examples and/or grammar and collocation information are provided. Except for MELD, the websites offer access to several different individual dictionaries. To imitate typical user behaviour, default settings for English language searches are used, and only those dictionaries from which data is presented on the initial results page are considered.



Stage 1: Analysis of coverage. Evaluating dictionary coverage involves not only judging if an item is covered, but also *how* it is covered. This study distinguishes between main entries and sub-entries. Across all the dictionary websites, in main entries the target word is listed as a headword. Sub-entries are more diverse. CAM and OX do not use sub-entries for derived forms. CD often lists derived forms as sub-entries as part of the main entry for the base form (Figure 1). On CD, derived forms are often, simultaneously, presented at the foot of the main entry for the base form under the heading "Derived forms" (Figure 2). In many LONG entries, "Word families" containing derived forms are shown at the top of the results page. On LONG, derived forms are sometimes presented as sub-entries (Figure 3). MELD often lists derived forms at the foot of the main entry for the base word under the heading "Derived word" (Figure 4). For ease of comparison, all these variations in sub-entry presentation are labelled 'sub-entry' here.

Definition of 'accessible'

accessible


Collins COBUILD


Word Frequency ●●●●●


(æksɛsɪbəl)  



1. **ADJECTIVE** **B2**


If a place or building is **accessible** to people, it is easy for them to reach it or get into it. If an object is **accessible**, it is easy to reach.

The Centre is easily accessible to the general public. [*ˌtɛntə*] 

The premises are wheelchair accessible. 

I kept my phone readily accessible in case I saw something that needed to be filmed. 

accessibility (æksɛsɪbɪlɪti)   **UNCOUNTABLE NOUN**

...the easy **accessibility** of the area. [*ˌɛsɪˈbɪlə*] 

Synonyms: openness, susceptibility, exposedness [More Synonyms of accessible](#)

Synonyms: friendliness, informality, cordiality, affability [More Synonyms of accessible](#)

Synonyms: approachability, availability, readiness, nearness [More Synonyms of accessible](#)

Figure 1: Entry for **accessible** from CD with derived form **accessibility** as sub-entry (highlighted)

Derived forms



accessibility (acˌcessɪˈbɪlɪti) **NOUN**

accessibly (acˌcessɪbly) **ADVERB**

accessible

Word Frequency ●●●●●

in American English

(æksɪˈsɛsəbəl ); (əkˈsɛsəbəl 

ADJECTIVE

1. that can be approached or entered
2. easy to approach or enter
3. that can be got; obtainable
4. open to the influence of *with to accessible to pity*
5. easily understood or generally appreciated
an accessible poet

Webster's New World College Dictionary, 4th Edition. Copyright © 2010 by Houghton Mifflin Harcourt. All rights reserved.

Derived forms

accessibility (acˌcessɪˈbɪlɪti) **NOUN**



accessibly (acˌcessɪbly) **ADVERB**

Figure 2: Entry for **accessible** from CD with derived form **accessibility** as listed under "Derived forms" (highlighted)


accessible


Word family (noun) access **accessibility** ≠ **inaccessibility** (adjective) accessible ≠ inaccessible (verb) access (adverb) accessibly ≠ inaccessibly

From Longman Dictionary of Contemporary English

ac-ces-si-ble /əkˈsesəbəl/ •○○ **AWL** adjective  


1 a place, building, or object that is accessible is easy to reach or get into **OPP** inaccessible

 The island is only accessible by boat.


 There is a church which is **easily accessible** from my home.

2 easy to obtain or use


accessible to

 the need for a health service that is accessible to all

easily/readily accessible


 Computers should be made readily accessible to teachers and pupils.

3 someone who is accessible is easy to meet and talk to, even if they are very important or powerful **SYN** approachable

 I think that you'll find she's very accessible.

4 a book, poem, painting etc that is accessible is easy to understand and enjoy

accessible to

 He wants his music to be accessible to everyone.

—accessibly **adverb**

—**accessibility** /əkˈsesəˈbɪləti/ **noun** [uncountable]

Figure 3: Entry for **accessible** from LONG with derived form **accessibility** as listed at end of main entry

DERIVED WORD

accessibility

NOUN

the quality and accessibility of materials

Figure 4: Derived form *accessibility* as listed at end of main entry for **accessible** from MELD

Additionally, the websites' response to searches for rare wordforms differs. CAM sometimes uses placeholder examples retrieved automatically from a corpus. If no standard entry can be found, MELD occasionally redirects the user to an example from its crowd-sourced *OPEN DICTIONARY*. In this analysis, placeholder and crowdsourced examples are treated as coverage provided examples are relevant to the target word. CD, LONG, MELD, and OX redirect the user to the more common wordform (e.g., *philosophic* redirects to *philosophical*). The common and rarer forms are considered interchangeable.

The treatment of words with ambiguous word classes, principally *-ed* affixes in the sample, varies within and between the dictionary websites. Searches for wordforms presumed adjectives by Schmitt and Zimmerman (2002) ending in *-ed* often redirect to the entry page for the verb in MELD and OX which use different pages for word classes, and to the main-entry page in covering both

noun and verb in the other resources. If a sub-entry exists on these pages for the adjectival form, this form is recorded as being covered (e.g., **authorize** in LBD). Occasionally, traditional examples or corpus lines (automatically generated, occasionally incomplete sentences) illustrate an adjectival use even though the adjectival sense is not explicitly covered (e.g., *release* in CAM: "To what extent the rural sector absorbs the *released labour* is not clear"). In these cases, the presence of a relevant example or corpus line is noted for stage two of the analysis.

Stage 2: Analysis of support for written production. A key assumption of this study is that examples and information about typical grammatical and collocational behaviour support productive use of the wordforms. While there is much research about what constitutes a good dictionary example (Kilgarriff et al. 2008) and the optimal number of examples for supporting production (Frankenberg-Garcia 2015; Ptasznik 2023), here analysis is limited to noting the presence or absence of examples.

In this study, typical combinations containing grammatical words (i.e., prepositions and determiners) are labelled grammar patterns while typical combinations of lexical words are labelled collocations. This policy is maintained irrespective of how these combinations are labelled on the dictionary websites. For example, combinations of grammatical words often appear in the collocation dictionary sections of the websites. The theoretical debate about the difference between collocation and grammar pattern is irrelevant for most dictionary users. However, information about the lexical items which co-occur with a particular wordform, and their syntactic configuration is useful for writers.

By aggregating the number of entries with writing support features such as examples, grammar patterns, and collocation information and dividing this by the total number of items from the sample covered, a *writing support score* can be calculated. This score gives an approximation of how well a resource supports users with the sample items when writing.

$$R = \frac{E_x + G_x + C_x}{N}$$

The overall writing support score is the ratio (R) of the sum of items with examples (E), grammar patterns (G), and collocations (C) for the sample items covered by the dictionary (N). To reflect the diversity in syntactic behaviour of word classes and the varying degrees of difficulty they could present writers, four writing support scores are reported:

- Overall score
- Score with adverbs excluded
- Score with adjectives excluded
- Score with adverbs and adjectives excluded

To calculate the exclusive scores, the sum for items of the included word classes (T) are first weighted (W) representing their proportion of the total sample items covered by the dictionary (N):

$$W_c = \frac{T_c}{N}$$

The exclusive score, a ratio, is then calculated using this weighting.

$$r = \frac{(T_x * W_c) + G_x + C_x}{N}$$

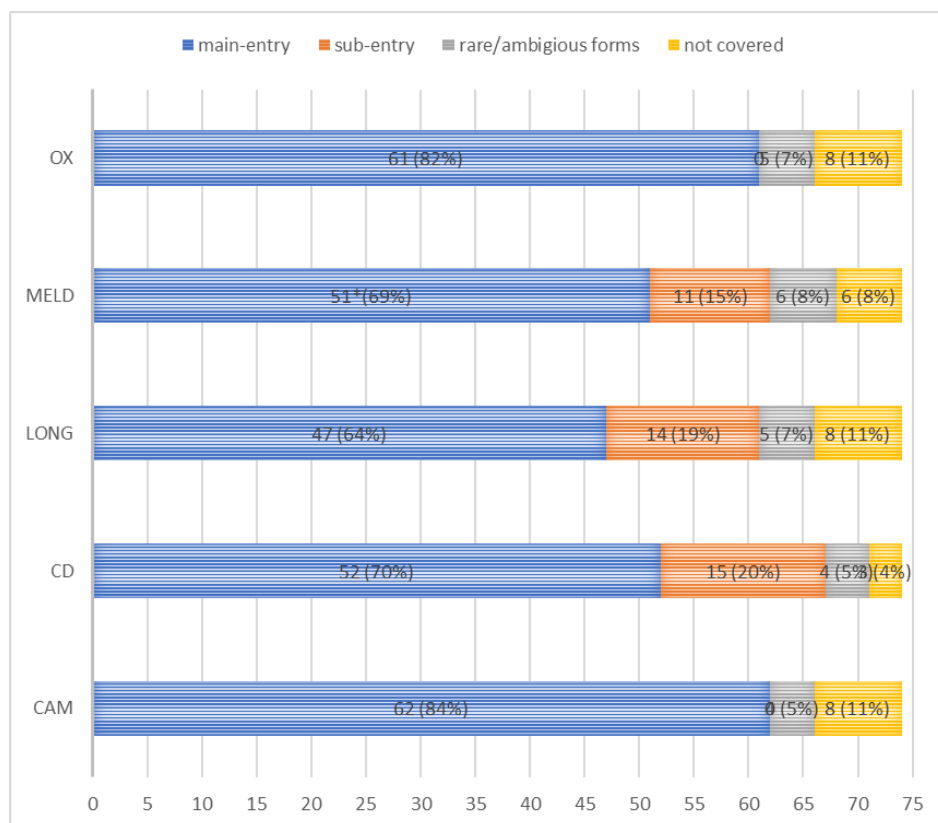
The coverage statistics and writing support scores indicate how well users of the dictionary websites are supported when seeking to use the problematic derivative forms in writing. For a more detailed impression, it is necessary to examine which items have writing support features.

Although examples can provide information about grammar patterns, here analysis focuses on semantics. Namely, whether derived wordforms missing examples are sufficiently semantically regular for a user to infer their meaning and use. This study does not differentiate between exemplification styles employed in the dictionaries. However, it is noteworthy that CAM, CD, and LONG occasionally present examples automatically extracted from corpus lines. When relevant to the target word, these are counted.

Comparing items with and without grammar pattern and collocation information by word class across the dictionary websites provides a clearer impression of how well users are supported when writing the problematic forms. Although users can intuit collocation and grammar patterns from examples, only those instances where the dictionary compiler intentionally highlights these aspects are considered. Common strategies include presenting salient collocations or grammar patterns in bold in examples (all dictionary websites examined) and/or separating common collocates with slashes (e.g., LONG (Figure 3) and OX) and displaying information from the publisher's collocation dictionary for certain searches. Additionally, LONG occasionally provides links to fuller entries for salient collocations and grammar patterns; CD, MELD, and OX display common idioms for some of the sample, while CD includes COBUILD grammar patterns.

Results and discussion

Coverage. The impression of inconsistent treatment of derivative forms reported in previous research is not immediately supported. Most items in the sample are covered by the five websites. The mean number of items treated per website ($N = 74$) is 66.4 with a standard deviation of 3.64. The overall coverage of the sample items did not differ significantly by dictionary website, $X^2 = 3.194$; $df = 4$; $p < .05$.



*Includes entry for Macmillan OPEN DICTIONARY for *persistency*

Figure 5: Coverage and entry status of problematic wordform on the "Big five" dictionary websites

A high degree of coverage was expected, the sampling criteria ensured target items were used reasonably frequently and widely. Indeed, inclusion in a dictionary was one of the three criteria Schmitt and Zimmerman (2002) used to select the permissible responses to their gapped sentence exercise.

Greater difference is apparent in *how* words are treated. OX and CAM cover all sample items as main entries, while CD, LONG, and MELD use sub entries for around one-sixth of the items. This suggests that although CAM and MELD cover a greater number of items overall, OX and CAM provide better writing support than the other resources. Further analysis of the entry contents is needed to substantiate this.

Table 4: Wordforms which are missing from at least one of the dictionary websites

	CAM	CD	LONG	MELD	OX
<i>accessibly</i> (Level 3)	x	sub	sub	x	x
<i>authoritive</i> (Level 6)	x	x	x	x	x
<i>authoritively</i> (Level 3)	x	x	x	x	x
<i>coherency</i> (Level 6)	x	redirect	redirect	x	x
<i>ethnicity</i> (Level 4)	main	main	x	main	main
<i>liberalness</i> (Level 3)	x	sub	x	x	x
<i>minimization</i> (Level 4)	main	sub	x	sub	main
<i>persistency</i> (Level 4)	x	redirect	x	open	
<i>philosophic</i> (Level 6)	placeholder	main	redirect	redirect	redirect
<i>preciseness</i> (Level 3)	main	sub	x	sub	x
<i>traditionize</i> (Level 4)	x	x	x	x	x

Table 4 shows the eleven wordforms which are missing from at least one website. Only three items are absent from all websites: *authoritive* and *authoritively*, infrequent spellings of *authoritative* and *authoritatively*, and *traditionize* a rarer verb meaning 'to make into a tradition'.

The treatment of *coherency*, a more infrequent form of *coherence*, and *philosophic*, a more infrequent form of *philosophical*, is inconsistent. The former is absent from CAM, MELD, and OX, the latter not found in CAM. Except for *philosophic* in CAM and CD, searching for these wordforms redirects the user to the page for the more frequent form. Once there, the infrequent form is listed after "also" (LONG and OX) or "or" (MELD). The first entry when searching for *philosophic* on CD is a COBUILD entry stating: "*Philosophic* means the same as *philosophical*" with a hyperlink to *philosophical*. Since both wordforms are wholly interchangeable, this redirection strategy seems sound. For resources where the infrequent forms are not listed, the alphabetic proximity of these items to their counterparts means that users may select the relevant form from the alphabetical listing presented when a search produces no exact results. Searching for *philosophic* in CAM produced a placeholder consisting solely of corpus lines for *philosophic*.

The treatment of forms with the Level 3 affix *-ness*, *liberalness* (only present in CD) and *preciseness* (absent from LONG and OX), may be inconsistent. For example, *preciseness* is in CAM but not *liberalness*. It may be that *liberalness* was considered too infrequent for inclusion⁶ or its inclusion may be an oversight given the productivity (almost any adjective + *-ness* produces an acceptable noun) and semantic regularity (meaning "'property of being X', where X is the base

adjective" (Carstairs-McCarthy 2018: 78)) of this suffix. However, as these are the only two *-ness* forms in the sample, care must be taken not to overgeneralise.

The omission from CAM, MELD, and OX of *accessibly*, an adverb formed with the Level 3 affix *-ly* could suggest inconsistent coverage. However, the presence of the thirteen other *-ly* adverbs from the sample suggests another factor, possibly frequency, plays a role.

The wordforms *ethnicity* and *minimization* are notably absent from LONG. There are four other occurrences of *-ity*, and two other occurrences of *-ation* sample wordforms covered by the website. Since words formed with *-ity* often have a specialised meaning which "may be hard to deduce" (Bauer and Nation 1993: 275), the omission of *ethnicity* is unfortunate. The omission of *minimization* here is surprising given its frequent semi-technical uses. While the omission of these words formed with often challenging Level 4 affixes could be a simple mistake, it may still inconvenience users.

Beyond coverage, there is less consistency in the way sample items are treated across the websites. One source of confusion is the ambiguous status of *-ed* and *-ing* forms which can be analysed as either adjectives or participle forms and in the case of *-ing* also as nouns. Schmitt and Zimmerman (2002) label the *-ed* forms (*accessed*, *assumed*, *authorized*, *released*, and *selected*) and the *-ing* form (*surviving*) as adjectives.

Table 5: Treatment of ambiguous word class forms

	CAM	CD	LONG	MELD	OX
<i>accessed</i>	verb	verb	verb	verb	verb
<i>assumed</i>	verb	adjective	verb	verb	adjective
<i>authorized</i>	adjective	adjective	adjective	verb	adjective
<i>released</i>	verb	verb	verb	verb	verb
<i>selected</i>	verb	verb	adjective	verb	verb
<i>surviving</i>	adjective	adjective	adjective	adjective	verb

Table 5 shows searches for these *-ed* and *-ing* forms give inconsistent results. All sites redirect searches for the items *accessed* and *released* to *access* and *release* (v). No adjectival senses of these items are given. The adjective sense of *selected* is a sub-entry of the verbal sense from the LBD. The adjective *assumed* is listed as a main entry in CD and OX. The adjective *authorized* is present as a main entry in all the dictionaries except MELD. The adjective *surviving* is present as an entry or sub-entry in all dictionaries except OX. However, there are examples and collocations for the verbal entry which could be analysed as adjectival.

Some of these deficiencies are mitigated, intentionally or otherwise, by fea-

tures of online dictionaries. Problems with corpus methods in lexicography often stem from inaccuracies in part-of-speech tagging (Frankenberg-García, Rees and Lew 2021). Many methods tend to treat *-ed* forms as verbs rather than adjectives. This may explain the tendency to treat these forms as participles in the dictionaries. However, it also means that some of the corpus-derived examples in verbal entries could be analysed as adjectives. For instance, the example provided for the fifth sense of *release* (v) in OX: "The newly *released files* reveal [...]". This is more apparent still in automatically retrieved examples from corpora. For example, in the entry for *release* in CAM: "To what extent the rural sector absorbs the *released labour* is not clear" and "There are only a few landraces and very old *released varieties* available."

Helpful features include the alphabetical index adjacent to entries on all websites except LONG. For example, on CAM's page for **assume**, the user is presented with adjectival uses: *assumed debt*, *assumed liabilities*, *assumed name* in the 'Browse' box at the bottom of the entry. For years, liberation from the constraints of the alphabetical index has been regarded positively (c.f., De Schryver 2003). However, this feature can mitigate a methodological deficiency in electronic lexicography. Predictive text searches also help users find adjectival senses. For example, in MELD typing *assumed* predicts *assumed name* which is listed as a discrete entry. MELD also contains a crowd-sourced example containing an adjectival use of *authorized*, *authorized push payment*. An example of a crowd-sourced element potentially resolving a deficiency, albeit a relatively minor one, in a professionally produced dictionary.

The prevalence of homographs in English is problematic for electronic lexicography. Table 6 indicates the word class initially displayed when searching for a homographic item. Dictionary search engines cannot determine the user's intended word class. The basic form's ordering might reflect the compilers' view of the primary form or merely the frequency of word classes in the corpora used.

Table 6: Word class first presented for homographic pair

	CAM	CD	LONG	MELD	OX
<i>access</i> (n) vs. <i>access</i> (v)	noun	noun	noun	verb	noun
<i>minimum</i> (n) vs. <i>minimum</i> (adj.)	noun	noun	adjective	adjective	adverb
<i>release</i> (n) vs. <i>release</i> (v.)	verb	verb	verb	verb	verb
<i>select</i> (v) vs. <i>select</i> (adj.)	verb	verb	verb	verb	verb

This coverage analysis provides insights into how members of derivationally related WFs are treated in online monolingual English dictionaries. Overall coverage statistics suggest reasonably consistent treatment of the WF members sampled. Inconsistencies include: the omission of forms with the morpheme *-ness* (*liberalness* and *preciseness*) which could be justified by its formal and semantic regularity, inconsistent treatment of rare wordforms which have more frequent

equivalents (*coherency* and *philosophic*), and the ambiguous word class of *-ed* and *-ing* wordforms. These minor inconsistencies may not have an impact on the user. Furthermore, electronic lexicography methods both contribute to and mitigate such inconsistencies.

Writing support. The coverage analysis above suggests that members of derivationally related WFs are well covered on the websites examined (RQ1). However, to establish the extent to which they are treated in a way which facilitates productive use in writing (RQ2) a finer-grained analysis is necessary. A key assumption here is that examples, grammar patterns, and collocation information help writers. Another assumption is that the six rare forms with more frequent counterparts can be disregarded since it is likely that users will look up the more frequent counterpart.

Table 7: Writing support scores

	CAM	CD	LONG	MELD	OX
Overall score	1.9	2.1	1.8	1.5	2.1
Score with adverbs excluded	2.2	2.1	1.9	1.8	2.3
Score with adjectives excluded	2.0	1.9	1.7	1.6	2.1
Score with adverbs and adjectives excluded	1.8	1.7	1.5	1.4	1.9

The writing support scores in Table 7 suggest that OX provides the most comprehensive writing support for the problematic wordforms, closely followed by CAM and CD. MELD's score is notably lower than the others. This relation holds for the exclusive scores. However, caution is needed when interpreting differences in such a small sample. These scores indicate inconsistency in *how* the sample is treated across the websites examined. A closer examination of the individual components of writing support (examples, grammar patterns, and collocation information) confirms this impression and elucidates differences in sample treatment within dictionary websites.

The proportion of items with examples (Figure 6) differs significantly by website, $X^2 = 30.068$; $df = 4$; $p < 0.001$. Both OX and CAM provide examples for 97% of the items they cover. Items missing examples are the ambiguous word class forms *accessed* (CAM and OX), *assumed* (CAM), and *selected* (OX). Since the dictionaries treat them as verbs and provide examples for the verbal senses, they effectively offer examples for all items they list. CD provides examples for 88% of items covered. Again, two ambiguous class items (*accessed* and *released*) lack examples. LONG provides examples for 86% of items it covers including *accessed*, *assumed*, and *released*. The outlier here is MELD where 70% of items covered have examples.

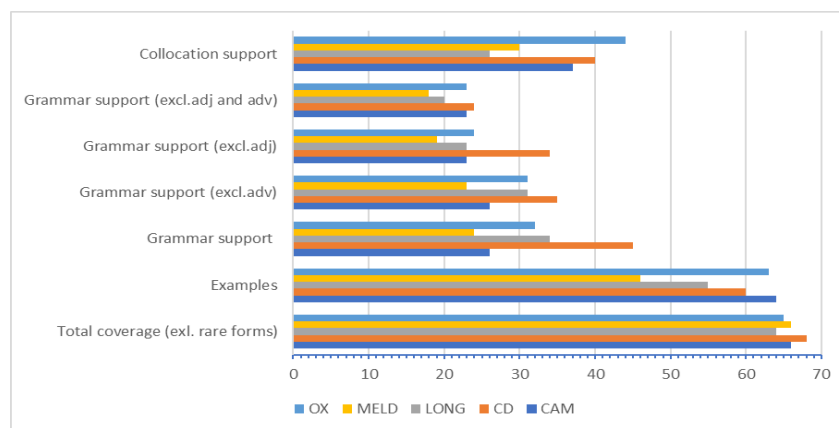


Figure 6: Sample items with writing support features per resource

There is clear inconsistency in the provision of grammar patterns on the websites examined (Figure 6). The proportion of items with this information differs significantly by website, $X^2 = 14.2796$; $df = 4$; $p < 0.006$. Overall, CD leads providing information for 60% of items covered. LONG provides grammar information for 53% of items covered, followed by OX (49%). CAM provides grammatical information for 39% of covered items, MELD for 36%. The syntactic behaviour of different parts-of-speech poses different degrees of challenge for writers. However, this trend persists when adverbs are excluded. For example, with a coverage statistic of 72% CD is notably higher than OX (55%), LONG (53%), and CAM (51%), and considerably more so than MELD (42%). When adjectives and adverbs are excluded, OX has the highest statistic (72%) followed by CD (71%) and CAM (70%); LONG covers 67% of noun and verb items, with MELD lower at 55%.

The proportion of items with collocation information (Figure 6) differs significantly by website, $X^2 = 12.192$; $df = 4$; $p < 0.05$. OX leads by providing collocation information for 68% of items covered, followed by CD (59%) and CAM (56%), then MELD (45%), and finally LONG (41%).

Since many users can induce information about grammatical patterns and collocational behaviour from dictionary examples and corpus lines, the relative absence of grammar patterns on CAM and OX is perhaps mitigated by their comprehensive example provision. This is reflected in the overall writing support score.

From the broad view adopted so far, considerable variation in the provision of writing support features between dictionaries is apparent. The following three sub-sections provide a finer-grained analysis of this variation.

Examples. As Table 8 indicates, after ambiguous word class forms, *-ly* adverbs are the wordforms most frequently missing examples. In general, they are semantically regular "Xly means 'in an X fashion' for any adjective X" (Carstairs-McCarthy 2018: 20). This general rule applies to *coherently*, *ethnically*,

minimally, and *persistently* (all lacking examples in MELD). However, *accessibly*, *authoritatively*, *ideologically*, and *philosophically* are edge cases. For instance, without an example learners lacking deep relational knowledge could conceivably make the erroneous connection *philosophy* → *philosophical* ('related to philosophy') → *philosophically* (in a 'manner related to philosophy') rather than the prototypical meaning "in a way that calmly accepts a difficult situation" (CAM).

Table 8: Items missing examples

	CAM	CD	LONG	MELD	OX
<i>accessed</i>	X	X	X	X	X
<i>accessibility</i>			X		
<i>accessibly</i>	NC	X	X	NC	NC
<i>assumed</i>	X		X	X	
<i>authoritatively</i>			X	X	
<i>authorized</i>				X	
<i>cohere</i>				X	
<i>coherence</i>				X	
<i>coherently</i>				X	
<i>ethnically</i>				X	
<i>ethnicity</i>			NC	X	
<i>ideologically</i>		X	X		
<i>liberality</i>				X	
<i>liberalization</i>				X	
<i>liberalize</i>				X	
<i>liberalness</i>	NC	X	NC	NC	NC
<i>minimally</i>				X	
<i>minimization</i>		X	NC	X	
<i>persistently</i>				X	
<i>philosophically</i>		X	X	X	
<i>philosophize</i>				X	
<i>preciseness</i>		X	NC	X	NC
<i>released</i>		X	X	X	
<i>selected</i>				X	X
<i>selectively</i>			X		

Note. NC = not covered by dictionary website; X = missing

The high degree of productivity and semantic regularity of the affix *-ness* which generally means "property of being X", where X is the base adjective." (Carstairs-McCarthy 2018: 78) could explain the omission of *liberalness* from all resources except CD and *preciseness* from LONG and OX, and the omission of an example for *liberalness* (CD) and *preciseness* (CD and MELD). However, the presence of examples for these items in the other resources suggest their creators do not share this assumption of relational knowledge.

The lack of examples for *liberalization* (MELD) and *minimization* (CD and MELD) can be explained by the generalizability of *-ation*. However, as with the absence of *minimization* from LONG, both wordforms have a specialised meaning frequent in academic context (e.g., "He is a longtime proponent of his country's economic liberalisation." (CAM); "cost minimization" (OX)). The absence of *liberalize* from MELD is notable for the same reason (e.g., "They will work with a view to further liberalize the investment regime" (CAM)). Like the absence of an example for *philosophically* discussed in the coverage analysis above, the absence of an example for *philosophize* from MELD is problematic as it does not typically mean 'to create philosophy' rather "to talk for a long time about subjects such as the meaning of life" (CAM). An example could also demonstrate that, in contrast to many words derived with the affix *-ize*, it is intransitive. The following examples from CAM for the entries for the *-ize* forms sampled illustrate complementation patterns well:

I authorized my bank to pay her £3,000.

They have plans to liberalize the prison system.

We must minimize the risk of infection.

Students, she complained, had nothing better to do than spend whole days philosophizing about the nature of truth.

The provision of examples for wordforms derived with *-ity* is also problematic. The missing example for *accessibility* in LONG is surprising. Firstly, because examples exist in the other dictionaries and, secondly, because it has a specialised yet frequent sense: "how easy something is to reach, enter, use, etc. for somebody with a disability" (OX). Additionally, examples for *ethnicity* and *liberality* are missing from MELD. As Schmitt and Zimmerman (2002) show, the extent to which productive knowledge of these words is easily predictable from productive knowledge about their base is questionable. When the base has two or more senses this assumption of relational knowledge entails a further assumption: that the user knows which sense is relevant to the derivative. For example, the definitions below come from CAM: (1) and (2) define *ethnic*, (3) defines *ethnicity*. The relation between (1) and (3) is immediately apparent. The relation between (2) and (3) requires some mental gymnastics.

1. *relating or belonging to a group of people who can be seen as distinct (= different) because they have a shared culture, tradition, language, history, etc.:*

2. *seen as different or interesting because of coming from a culture or tradition that is not Western:*
3. *a large group of people with a shared culture, language, history, set of traditions, etc., or the fact of belonging to one of these groups:*

Examples for the WF members *coherence* and *cohere* are notable omissions from MELD. This may stem from an assumption that learners have the relational knowledge to make the connection to the adjective *coherent*. This is particularly questionable in the case of *coherence* as although the Level 5 affix *-ence* is reasonably regular, it is not frequent (Bauer and Nation 1993: 260).

Regarding the provision of examples, there is clear inconsistent treatment across dictionaries, and in the case of CD, LONG and MELD, within dictionaries. Barring the ambiguous word class items, in CD and LONG the sub-entry status of items may be an explanatory factor for, or a consequence of, the missing examples. However, in MELD both main- and sub-entries lack examples.

Grammar patterns. A comparison of items with (Figure 7) and without (Figure 8) grammar pattern information suggests inconsistent writing support between and within websites.

As discussed, the need for grammar pattern information varies by word class. All sampled adverbs given grammatical support in CD have main entry status. Their grammar patterns come from COBUILD. Many first appeared in the 'extra-column' of the paper dictionary (Hands 2018) and were migrated online. Wordforms lacking grammar pattern support occur as "derived words" and sub-entries in other Collins dictionaries such as *Collins English Dictionary* and *Webster's New World College Dictionary*. Similarly, entry status explains the presence of grammatical information for adverbs in LONG. Those with support are the "Sentence adverbs", *inevitably* and *traditionally*, and *precisely*. The latter is followed by the interrogative pronouns *how/when/where*. All adverbs lacking grammar support in LONG, except *liberally*, are sub-entries. In MELD, *precisely* is also listed followed by *how/when/what* and in OX it followed by *because*. All other sampled adverbs in the latter two dictionaries lack grammatical pattern information.

Adjectives selecting prepositions (*accessible to*, *liberal with*, *minimum of* etc.) are treated fairly consistently. Inconsistencies occur in CD, LONG, and OX, which mark typical word order for some adjectives (e.g., "precise [adj NOUN]") but not others with the same order (e.g., *coherent*). Dictionaries that do not indicate this order (e.g., CAM and MELD) offer less detailed yet more consistent treatment.

All sampled verbs in CD have grammar pattern information. Patterns for *cohere* are absent from CAM, LONG, and MELD. Of the *-ize* affixed verbs, only *liberalize* has pattern information in CD, while patterns for *philosophize* are absent in LONG, MELD, and OX. As discussed, grammar pattern information may be useful for learners wishing to use *philosophize* as it is a rare example of an intransitive verb derived with *-ize* which frequently occurs with the prepositions *of* or *about*, as documented in CAM and CD. Similarly, *cohere with* is a typical pattern given in OX and CD.

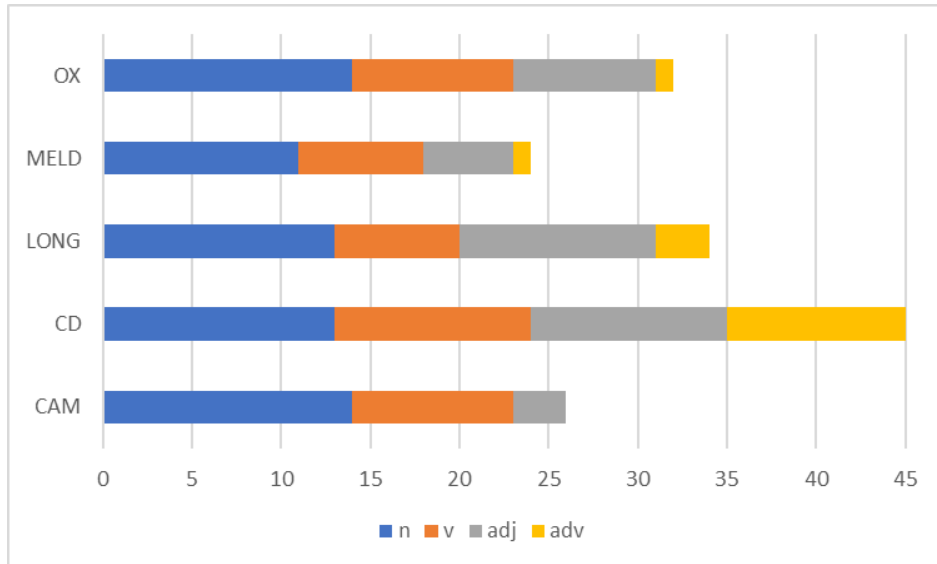


Figure 7: Sample items with grammar pattern information per resource

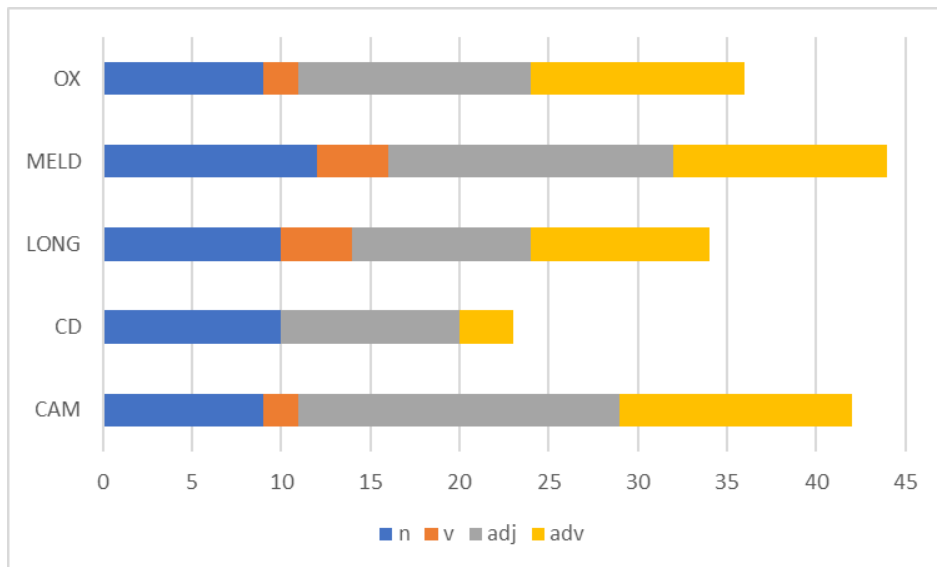


Figure 8: Sample items without grammar pattern information per resource

Nouns are derived using a greater variety of affixes than other word classes. Table 9 shows the sample nouns included on each website and whether they have grammar pattern information. The overall impression is one of inconsistent treatment within and between dictionaries.

Table 9: Nouns with grammar pattern information

	Without grammar patterns	Words with grammar patterns
CAM	<i>ethnicity</i> <i>ideology</i> <i>inevitability</i> <i>liberality</i> <i>liberalization</i> <i>preciseness</i> <i>precision</i> <i>survival*</i> (The phrase <i>survival of the fittest</i> is listed)	<i>access</i> <i>accessibility</i> <i>assumption</i> <i>authority</i> <i>authorization</i> <i>coherence</i> <i>liberty</i> <i>minimization</i> <i>minimize</i> <i>minimum</i> <i>persistence</i> <i>philosophy</i> <i>release</i> <i>selection</i> <i>tradition</i>
CD	<i>authority</i> <i>authorization</i> <i>coherence</i> <i>ethnicity</i> <i>ideology</i> <i>liberality</i> <i>liberalness (sub)</i> <i>minimization (sub)</i> <i>preciseness (sub)</i> <i>survival*</i> (The phrase <i>survival of the fittest</i> is listed)	<i>access</i> <i>accessibility</i> <i>assumption</i> <i>inevitability</i> <i>liberalization</i> <i>liberty</i> <i>minimize</i> <i>minimum</i> <i>persistence</i> <i>philosophy</i> <i>precision</i> <i>release</i> <i>selection</i> <i>tradition</i>
LONG	<i>accessibility (sub)</i> <i>authorization</i> <i>coherence</i> <i>ideology</i> <i>liberality</i> <i>liberalization (sub)</i>	<i>access</i> <i>assumption</i> <i>authority</i> <i>inevitability</i> <i>liberty</i> <i>minimum</i>

		<p><i>persistence</i> <i>philosophy</i> <i>precision</i> <i>release</i> <i>selection</i> <i>survival</i> <i>tradition</i></p>
MELD	<p><i>accessibility</i> (sub) <i>coherence</i> <i>ethnicity</i> <i>ideology</i> <i>liberality</i> <i>liberalization</i> (sub) <i>liberalness</i> <i>minimization</i> (sub) <i>minimize</i> <i>persistence</i> <i>philosophy</i> <i>preciseness</i> (sub) <i>precision</i></p>	<p><i>access</i> <i>assumption</i> <i>authority</i> <i>authorization</i> <i>inevitability</i> <i>liberty</i> <i>minimum</i> <i>release</i> <i>selection</i> <i>survival</i> <i>tradition</i></p>
OX	<p><i>ethnicity</i> <i>ideology</i> <i>inevitability</i> <i>liberality</i> <i>liberalization</i> <i>minimization</i> <i>persistence</i></p>	<p><i>access</i> (noun) <i>accessibility</i> <i>assumption</i> <i>authority</i> <i>authorization</i> <i>coherence</i> <i>liberty</i> <i>minimize</i> <i>minimum</i> <i>philosophy</i> <i>precision</i> <i>release</i> <i>selection</i> <i>survival</i> <i>tradition</i></p>

Wordforms without grammar pattern information are predominantly derived by affixation using *-ity* and *-ation*. Those that do have grammar patterns can be analysed as the base wordforms or are often the most frequent member of their family according to Schmitt and Zimmerman's (2002) counts. The usefulness of grammar pattern information for these items to writers can only be ascertained by direct empirical research. However, it is notable that producing these wordforms posed problems for Schmitt and Zimmerman's (2002) participants.

Some items missing grammatical patterns exhibit similar grammatical

behaviour to those which have them. For example, *assumption that* appears in all resources while *inevitability that* is absent from CAM and MELD. This suggests a need for grammatical pattern information for many items missing it. Like the provision of examples, many of the wordforms without grammatical pattern information were treated as subentries, irrespective of their word class.

Collocation information. Unlike closed classes or phrasal categories that constitute grammar patterns, the range of potential collocates is limitless. Variation in typical collocates presented for a given base between resources is expected due to variation in corpus composition. Consequently, this analysis of collocation information must adopt a broad focus.

The provision of collocation information does not follow the general trend for writing support in the dictionaries examined. Notably, LONG rather than MELD provides collocation information for fewest items. However, differences exist across word classes.

Table 10: Noun items missing collocation information

	CAM	CD	LONG	MELD	OX
<i>accessibility</i>		X	X	X	X
<i>authorization</i>			X		
<i>coherence</i>			X		
<i>ethnicity</i>	X	X	NC	X	X
<i>ideology</i>			X		
<i>inevitability</i>	X			X	X
<i>liberality</i>		X	X	X	X
<i>liberalization</i>		X	X	X	X
<i>liberalness</i>	X	X	NC	X	NC
<i>minimization</i>		X	NC	X	X
<i>minimum</i>		X			
<i>persistence</i>	X			X	
<i>philosophy</i>				X	
<i>preciseness</i>	X	X	NC	X	NC
<i>precision</i>				X	
<i>selection</i>		X			

Note. NC = not covered by dictionary website; X = missing

Collocation information is absent for three out of twenty-three noun items (Table 10) in all resources: *ethnicity*, *liberalness*, and *preciseness*. Five items (*accessibility*, *liberality*, *liberalization*, and *minimization* in CAM; and *inevitability* in COD)

only have it in one resource. In contrast, seven items are absent from one resource (*minimum* and *selection* from COD; *authorization*, *coherence*, *ideology* from LONG; and *philosophy* and *precision* from MELD).

Table 11: Verb items missing collocation information

	CAM	CD	LONG	MELD	OX
<i>access</i>	X			X	
<i>assume</i>	X				
<i>authorize</i>			X	X	
<i>cohere</i>		X	X	X	X
<i>liberalize</i>		X	X	X	
<i>minimize</i>	X	X			X
<i>persist</i>			X		
<i>philosophize</i>		X	X	X	X
<i>release</i>	X				
<i>select</i>	X		X		
<i>survive</i>			X		

Note. X = missing

All eleven verb items have collocation information in at least one resource (Table 11). Although, for *cohere* and *philosophize*, this information is only provided by CAM. This is problematic because it assumes relational knowledge with other family members. Three resources lack information for *liberalize* (CD, LONG, and MELD) and *minimize* (CAM, CD, and OX). As with examples, some academic writers might benefit from collocation information about these semi-technical terms.

Table 12: Adjective items missing collocation information

	CAM	CD	LONG	MELD	OX
<i>accessed</i>	X	X	X	X	X
<i>assumed</i>	X		X	X	
<i>authoritative</i>			X	X	X
<i>authorized</i>	X		X	X	X
<i>coherent</i>	X		X		
<i>ideological</i>	X		X	X	

<i>liberal</i>				X	
<i>minimal</i>			X		
<i>minimum</i>		X			
<i>philosophical</i>			X	X	
<i>precise</i>			X		
<i>released</i>	X	X	X	X	
<i>select</i>				X	
<i>selected</i>	X	X	X	X	
<i>selective</i>			X		
<i>surviving</i>	X	X		X	

Note. X = missing

Ostensibly, provision of collocation information for adjectives is less comprehensive than for nouns and verbs (Table 12). However, seven of the items missing collocation information are ambiguous word class items treated as verbs. Moreover, three resources lack information for *ideological* (CAM, LONG, and MELD) three for *authoritative* (LONG, MELD, and OX), two for *coherent* (CAM and LONG) and *philosophical* (LONG and MELD).

Table 13: Adverb items missing collocation information

	CAM	CD	LONG	MELD	OX
<i>accessibly</i>	NC	X	X	NC	X
<i>authoritatively</i>	X	X	X	X	
<i>coherently</i>	X	X	X	X	X
<i>ethnically</i>	X	X	X		X
<i>ideologically</i>	X	X	X	X	X
<i>inevitably</i>	X	X	X	X	X
<i>liberally</i>	X		X	X	X
<i>minimally</i>		X	X	X	X
<i>persistently</i>	X		X	X	X
<i>philosophically</i>	X	X	X	X	X
<i>precisely</i>	X				
<i>selectively</i>	X	X	X	X	
<i>traditionally</i>	X	X	X	X	X

Note. NC = not covered by dictionary website; X = missing

Provision of collocation information for adverbs is the least comprehensive of all word classes (Table 13). Information is provided for *precisely* in all resources except CAM. CD also provides information for *liberally* and *persistently*, OX for *authoritatively* and *selectively*, and CAM for *minimally*. Two factors may explain this sparse coverage: Firstly, the suffix *-ly* is extremely semantically regular "Xly means 'in an X fashion', for any adjective X." (Carstairs-McCarthy 2018: 20), so presumably lexicographers assume users can use the *-ly* adverbs in production by connecting them to their knowledge of the adjective base. Secondly, users are unlikely to start a collocation search using an adverb: "It would not make sense for a writer to initiate a collocation query from an adverb (e.g. 'what words can I use with *primarily*?')" (Frankenberg-Garcia et al. 2019: 28).

This analysis of grammar support features for derivative forms suggests examples, grammar patterns, and collocations work independently when supporting writers. This is unlikely; writers may take information simultaneously from all three sources. If one feature (e.g., grammar pattern information) is unavailable they may rely more heavily on another (e.g., examples). Future analysis of writing support would benefit from a model reflecting this relationship.

Conclusions

This study aimed to investigate the treatment of academic WFs on five English dictionary websites frequently used by learners. It was motivated by a belief that the members of these WFs should be treated in a way that facilitates learners' written production. Two factors prompted this belief: Firstly, research demonstrating that when given a basic prompt wordform, academic writers struggle producing derivative forms from the same WF. Secondly, the removal of space restraints in electronic resources, which hypothetically allows more detailed coverage of derivatives than paper-based dictionaries.

Overall, the five websites examined cover most items in the sample of challenging wordforms. This good coverage contrasts with findings on paper-based dictionaries. However, as in previous research, there is considerable variation in the treatment of derivative wordforms within and between resources.

The quantity of writing support features varies greatly across websites. Although MELD covers a high proportion of sample items, it provides fewer examples, grammar patterns, and collocation information than the other resources. Within resources, the reasons for inclusion or exclusion of items and their related writing support features are not always clear. For certain affixes, this may be due to assumptions about generalisability of their semantic or syntactic behaviour. These assumptions may be misguided since empirical research suggests writers do not always connect bases and derivatives formed by suffixation even with highly generalisable and productive affixes. Occasionally, (e.g., *ethnicity*, *liberalization*), analysis of the excluded wordforms suggests their semantic relationship to the base is idiosyncratic. Alternatively, their relative frequency in corpora used in compilation may explain exclusion. Further investigation here would be beneficial.

Further research could also mitigate limitations restricting the generalisability of these conclusions. Important limitations relate to the 74 problematic word-forms investigated. Not only is this sample small, but its items are also morphologically limited containing a relatively narrow range of suffixes. Future research should investigate forms created via prefixation (e.g., with *co-*, *in-*, *re-* etc.) if producing these is found to be a problem for writers.

Practical considerations for dictionary makers. Assumptions about users' relational knowledge of WF members should be reevaluated. Instead of assuming that writers can connect the base, the affix and derivative meaning, dictionary makers should aim for more complete treatment of derivatives. Electronic resources, unrestrained by the physical restrictions of paper-based dictionaries, could offer users fuller entries for derivative forms. However, compiling dictionary entries costs money. Deprived of income from sales of paper dictionaries, it is unlikely that publishers will invest in this. Nonetheless, as seen with corpus lines and collocation lists, methods from electronic lexicography can, sometimes inadvertently, offer a solution.

Endnotes

1. For a more nuanced view, see Lew (2011) who makes a distinction between the potentially unlimited *storage space* for lexicographic data and more limited *presentation space* on the user's screen.
2. *Macmillan English Dictionary online* was shut down on June 30th, 2023.
3. The edition of the LBD from which the entry is taken is not specified.
4. The editions of COBUILD and CED from which the entries are taken are not specified.
5. The dictionaries mentioned are *Cambridge International Dictionary of English* (Procter 1995), *COBUILD English Learner's Dictionary* (Sinclair 1989), *Longman Dictionary of English Language and Culture* (Summers 1992), and *Oxford Advanced Learner's Dictionary of Current English* (Crowther 1995).
6. This would be surprising; "we checked the frequency of these derivatives in the BNC and considered eliminating those that had very low frequency counts or did not exist in the corpus." (Schmitt and Zimmerman 2002: 156)

References

- Agnes, M. (Ed.). 2010. *Webster's New World College Dictionary*. Fourth Edition. Cleveland, Ohio: Wiley.
- Bauer, L. and P. Nation. 1993. Word Families. *International Journal of Lexicography* 6(4): 253-279.
- Berko, J. 1958. The Child's Learning of English Morphology. *Word* 14(2-3): 150-177.
- Carstairs-McCarthy, A. 2018. *An Introduction to English Morphology: Words and Their Structure*. Kindle Edition. Edinburgh: Edinburgh University Press.
- Coxhead, A. 2000. A New Academic Word list. *TESOL Quarterly* 34(2): 213-238.
- Crowther, J. (Ed.). 1995. *Oxford Advanced Learner's Dictionary of Current English*. Fifth Edition. Oxford: Oxford University Press.

- De Caluwe, J. and J. Taldeman.** 2003. Morphology in Dictionaries. Van Sterkenburg, P. (Ed.). 2003. *A Practical Guide to Lexicography*: 114-126. Amsterdam: John Benjamins.
- De Schryver, G.-M.** 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.
- Frankenberg-Garcia, A.** 2015. Dictionaries and Encoding Examples to Support Language Production. *International Journal of Lexicography* 28(4): 490-512.
- Frankenberg-Garcia, A., R. Lew, J.C. Roberts, G.P. Rees, and N. Sharma.** 2019. Developing a Writing Assistant to Help EAP Writers with Collocations in Real Time. *ReCALL* 31(1): 23-39.
- Frankenberg-Garcia, A., G.P. Rees and R. Lew.** 2021. Slipping Through the Cracks in e-Lexicography. *International Journal of Lexicography* 34(2): 206-234.
- Hands, P.** 2018. COBUILD Design and Layout: Changes over the Last 30 Years. *Collins Dictionary Language Blog*.
<https://blog.collinsdictionary.com/language-lovers/cobuild-design-and-layout-changes-over-the-last-30-years/> [30 July 2023]
- Hanks, P.** 2013. *Lexical Analysis: Norms and Exploitations*. Cambridge, MA: MIT Press.
- Hyland, K. and Polly Tse.** 2007. Is There an "Academic Vocabulary"? *TESOL Quarterly* 41: 235-253.
- Kilgarriff, Adam, Miloš Husák, Katy McAdam, Michael Rundell and Pavel Rychlý.** 2008. GDEX: Automatically Finding Good Dictionary Examples in a Corpus. Bernal, Elisenda and Janet DeCesaris (Eds.). 2008. *Proceedings of the 13th EURALEX International Congress, Barcelona, 15-19 July 2008*: 425-432. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- Lew, R.** 2011. Space Restrictions in Paper and Electronic Dictionaries and their Implications for the Design of Production Dictionaries. Bański, Piotr and Beata Wójtowicz (Eds.). 2011. *Issues in Modern Lexicography*. München: Lincom Europa.
- Procter, P. (Ed.).** 1995. *Cambridge International Dictionary of English*. Cambridge: Cambridge University Press.
- Ptasznik, B.** 2023. More Examples May Benefit Dictionary Users. *International Journal of Lexicography* 36(1): 29-55.
- Rees, G.P.** 2021. Discipline-Specific Academic Phraseology: Corpus Evidence and Potential Applications. Charles, M. and A. Frankenberg-Garcia (Eds.). 2021. *Corpora in ESP/EAP Writing Instruction: Preparation, Exploitation, Analysis*: 32-54. London: Routledge.
- Schmitt, N.** 1998. Tracking the Incremental Acquisition of Second Language Vocabulary: A Longitudinal Study. *Language Learning* 48(2): 281-317.
- Schmitt, N.** 1999. The Relationship between TOEFL Vocabulary Items and Meaning, Association, Collocation and Word-class Knowledge. *Language Testing* 16(2): 189-216.
- Schmitt, N. and P. Meara.** 1997. Researching Vocabulary through a Word Knowledge Framework: Word Associations and Verbal Suffixes. *Studies in Second Language Acquisition* 19(1): 17-36.
- Schmitt, N. and C. Zimmerman.** 2002. Derivative Word Forms: What Do Learners Know? *TESOL Quarterly* 36(2): 145-171.
- Sinclair, J. (Ed.).** 1989. *COBUILD English Learner's Dictionary*. London: Collins.
- Stein, G.** 1985. Word-formation in Modern English Dictionaries. Ilson, R. (Ed.). 1985. *Dictionaries, Lexicography and Language Learning*: 35-44. Oxford: Pergamon.
- Summers, D. (Ed.).** 1992. *Longman Dictionary of English Language and Culture*. Harlow: Longman.

- Ten Hacken, P., A. Abel and J. Knapp.** 2006. Word Formation in an Electronic Learners' Dictionary: ELDIT. *International Journal of Lexicography* 19(3): 243-256.
- Tyler, A. and W. Nagy.** 1989. The Acquisition of English Derivational Morphology. *Journal of Memory and Language* 28(6): 649-667.
- Van Sterkenburg, P.** 1992. *Het Woordenboek der Nederlandsche taal : Portret van een taalmonument*. The Hague: Sdu.
- West, M.** 1953. *A General Service List of English Words*. London: Longman, Green & Co.