Dominic Stewart

Grammar Labels for Verbs in English Monolingual Learners' Dictionaries

Abstract

The labels 'passive' and 'progressive,' as well as 'not passive' and 'not progressive,' are those most commonly adopted in English monolingual learners' dictionaries to indicate the grammatical category of verb headwords. However, it can happen that for specific verbs very different indications are provided from one dictionary to the next, a fact which would appear to derive primarily from diverging interpretations of corpus data on the part of lexicographers, and more specifically, from diverging interpretations of which corpus occurrences qualify as passive and progressive respectively for any given verb. This will lead to the discussion of a suggested conflict of form and function in corpus lexicography. Further, it is striking that the labels passive and progressive are prioritised at the expense of other verb labels such as imperative–used very sparingly in dictionaries–simple present, perfective and first person, which are not used at all. The corpus consulted is primarily the British Web 2007 (ukWac), but comparisons will be made with data from the British National Corpus.

Keywords: *lexicography, corpus linguistics, learners' dictionaries, frequency of verbs, grammatical categories, passive*

1. Introduction

Whether as teachers or as learners of English, all of us will have noticed that as a rule English monolingual learners' dictionaries provide labels such as 'formal,' 'informal,' 'dated,' 'usually negative,' etc. within word entries. Labels pertaining to grammar—generally linked to the word class (noun, adjective, verb etc.) of the headword—are also included, such as 'countable,' 'uncountable,' 'usually plural' for nouns, 'transitive,' 'intransitive,' 'usually passive' and 'usually progressive' for verbs, and 'not before noun' for adjectives. These grammar labels relate first and foremost either to the form of the headword itself, i.e., countable, uncountable, passive, progressive, imperative, or to what happens around the headword, i.e., transitive, intransitive, not before noun.

Although it could be argued that the most important and most consistent grammar information supplied in learners' dictionaries is of a colligational nature, i.e., connecting with the local

grammar of the entries (Dziemianko 2006), in this paper I intend to focus on those grammar labels which relate to the form of the headword, in particular those describing verbs, whose grammar is notoriously complex compared with other word classes (Jackson 2002, 108-109). As underlined by Bogaards and Van der Kloot (2001, 98), this focus is not uncharacteristic:

> In the discussions about grammatical information in learners' dictionaries, the syntax of the verb has always occupied the largest part. Although topics such as countable and uncountable nouns or gradable and non-gradable adjectives have also been taken into account, most of the attention and most of the energy have been directed towards questions concerning the verb.

However, this energy has been channelled into local grammar rather than into labels for headwords, which have received comparatively little attention. Rundell (2008a, 203-206) alludes to criteria that might justify or not justify a headword grammar label in the dictionary, distinguishing 'frequent passivisation,' whereby the verb in question tends to be defined in the active but is accompanied by the label 'often passive,' from 'dominant passivisation,' whereby the definition itself is in the passive, for example the verbs *apprentice* and *cheer* ("if someone is apprenticed to another person…;" "if you are cheered by something such as a piece of news…"). However, Rundell makes no reference to what the cut-off point might be between frequent and dominant passivisation, or indeed to what usage the term passivisation embraces. Bogaards and Van der Kloot (2001) devote an interesting paper to the use of grammatical information in learners' dictionaries, but are more concerned with the way entries are presented and with the range of grammatical codes adopted than with the use of headword labels.

This paper aims to redress the balance somewhat by focusing on the labels 'passive' and 'progressive,' firstly by debating what these labels actually mean, and secondly by questioning the wisdom of prioritising such labels at the expense of others. Issues of this nature are important because such is the success of English learners' dictionaries, a primary source of reference for millions of learners of English throughout the world, that one can easily become blinded to their potential flaws, especially since the principal beneficiaries of such dictionaries—learners of English—are unlikely to provide feedback on their user experience. It is clearly in the users' interests that the information contained in dictionaries is as consistent and unequivocal as possible.

The (online) dictionaries examined are the *Oxford Advanced Learner's Dictionary* (*OAL*), the *Macmillan Dictionary* (*Macmillan*), the *Longman Dictionary of Contemporary English* (*LDCE*), the *Collins Cobuild Reverso Dictionary* (*Cobuild*), and the *Cambridge Learner's Dictionary* (*CLD*). The corpus adopted is primarily the *British Web 2007*, but comparisons will be made

with data from the *British National Corpus*.[1] The frequency percentages supplied are based on random 200-line concordances extracted from the corpora. Unless otherwise stated, examples are drawn from *British Web 2007*.

2. The labels 'passive' and 'progressive' in English monolingual learners' dictionaries

In learners' dictionaries we sporadically come across verb labels such as 'often passive,' 'not passive,' 'usually progressive' and 'never progressive.' As pointed out by Atkins and Rundell (2008, 221):

> For verbs, information may be given about whether the headword is an activity, accomplishment, achievement, or stative verb, such as LDOCE's [*Longman Dictionary of Contemporary English*] 'not in progressive.'

The presence of these labels is sporadic for a number of reasons. Firstly, the bulk of grammatical information provided in English learners' dictionaries is collocational and colligational in nature, concerned above all with what happens adjacent to the headword but with less emphasis on the forms of the (lemma) headword itself. Secondly, and with specific regard to verbs, only a comparatively restricted number of verbs are marked for passive and progressive. Thirdly, the inclusion/exclusion of these labels varies significantly from one dictionary to the next. The *OAL* includes them relatively frequently, whereas more sparing use is made by the *Macmillan*, *Cobuild* and *CLD*. At the other end of the spectrum, the *LDCE* includes verb labels more rarely but prioritises "pattern illustrations" (Atkins and Rundell 2008, 401; Rundell 2008a, 199). For example, in the *LDCE* the verb *allege* is not explicitly flagged as often or usually passive, but two of the three patterns highlighted are *it is alleged* (*that*) and *be alleged to be*/*do something*:

> al·lege /əˈledʒ/ ●●○ **verb [transitive]** *formal*
> to say that something is true or that someone has done something wrong, although it has not been proved

[1] The *British Web 2007*, also known as *ukWac*, is a web-derived corpus assembled in 2007 containing over 1 billion 300 million words from websites in the .uk Internet domain. It is a general-purpose corpus with a broad range of text types. The *British National Corpus* contains approximately 100 million words of British English from the late twentieth century. It too is a general-purpose corpus offering a broad range of text types. It contains 90% written texts and 10% spoken. The two corpora have been consulted via The Sketch Engine, a corpus manager and analysis software created by Lexical Computing Ltd in 2003, now with over 500 corpora in more than 90 languages. See https://www.sketchengine.eu (Last visited 02/07/20) for further details.

it is alleged (that)
It was alleged that the policeman had accepted bribes.
allege that
The prosecution alleged that the man had been responsible for an act of terrorism.
be alleged to be/do something
The water is alleged to be polluted with mercury.

2.1 The labels 'often passive'/'usually passive'

Monolingual learners' dictionaries, all corpus-based, have made significant advances in the area of word frequency. The *CLD,* for example, specifies for each entry the corresponding level of the Common European Framework of Reference for Languages, while the *Macmillan* provides the appropriate number of frequency stars. See in particular Kilgarriff (1997) and Hanks (2012) on the crucial relationship between corpus-based frequency and lexicography. However, the use of frequency counts with reference to word classes (rather than lemmas) is rather more complex and has received barely a mention in the literature. With this in mind I shall now consider the dictionary labels 'often passive' and 'usually passive.'

Many grammars of English, for example the *Longman Grammar of Spoken and Written English* (Biber 1999), the *Cobuild English Grammar* (Hands 2017) and *Practical English Usage* (Swan 2017), provide a list of verbs often occurring in the passive. Some of the verbs they list are included in Table 1, which concerns the degree to which the passive label is present for certain verbs in monolingual learners' dictionaries (except the *LDCE*, for the reasons outlined above):

	acclaim	*paralyse*	*couch*	*inundate*	*entitle*	*allege*	*hamstring*	*prize*	*picture*
OAL	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Macmillan	Yes	Yes	Yes	No	Yes	No	no entry	No	No
Cobuild	Yes	No	Yes	Yes	Yes	No	No	Yes	Yes
CLD	Yes	Yes	No	No	Yes	Yes	No	Yes	Yes

Tab. 1: The presence of the labels 'often passive'/'usually passive' in English monolingual learners' dictionaries

Table 1 reveals that there are important differences from one dictionary to the next. Of the verbs chosen, only *acclaim* and *entitle* enjoy unanimous agreement among the compilers of the various dictionaries. This could to some extent be due to the fact that lexicographers use different corpora or have varying statistical methods of interrogating the data (see Rundell 2008a, 199 for a brief discussion of this issue with reference to the verb *bundle*), but it seems more plausible

that the differing outcomes are at least in part the result of differing interpretations of what constitutes a passive structure.

2.1.1 What does the label 'passive' mean?

Descriptions of the passive in grammars of English clarify unequivocally that a passive structure consists of the formula '*BE* + verb-*ed*,' thus for instance *she was acclaimed, they have been acclaimed, you will be acclaimed, he is being acclaimed*. All the examples supplied in the sections on the passive in the *Cobuild English Grammar,* in *Practical English Usage* and in the *Longman Grammar of Spoken and Written English* correspond to this formula. Now this might seem logical enough, but it needs to be underlined that '*BE* + verb-*ed*' excludes uses of, for instance, the verb *acclaim* such as the following, taken from the *British Web 2007* corpus.

(1) **Acclaimed** by so many as the ultimate villa holiday destination, the Algarve is justifiably renowned for its outstanding climate…

(2) Akira Kurosawa was the greatest of Japanese directors, **acclaimed** by Steven Spielberg as 'the pictorial Shakespeare of our time.'

(3) **Acclaimed** worldwide for the quality of its teaching and research, the University of Leeds is also the favourite amongst students applying for undergraduate courses.

(4) Rylance, **acclaimed** as one of Britain's most gifted young actors, is the self-effacing Jacob von Gunten.

Despite the absence of *BE* accompanying the past participle, all these occurrences of *acclaim* are characterised by what has traditionally been considered passive meaning: 1 and 2 feature the agent *by*, while in 3 and 4 the passive structure is implicit—*being acclaimed* or *having been acclaimed* in 3 and *who has been acclaimed* in 4. If these structures are not considered passive by lexicographers, then one might wonder if the baby is being thrown out with the bathwater, but whatever interpretation is intended, the dictionary user needs to be apprised of it.

Indeed, if only the prototypical '*BE* + verb-*ed*' is admitted as an eligible passive, then some of the labels relating to the passive found in dictionaries would appear to be unjustified. According to the *OAL*, *Macmillan*, *Cobuild* and *CLD*, the verb *couple* is either often or usually passive, but in both the *British Web 2007* and the *British National Corpus* the verb lemma query for *couple* retrieves the '*BE* + verb-*ed*' construction just over 10% of the time, a percentage which is far too small to warrant the descriptor 'usually passive.' Compare, for instance, the lemma query *inundate*, which in the *British Web 2007* and *British National Corpus* captures the canonical

passive structure at a rate exceeding 70%, and the verb lemma query *couch*, which in the two corpora retrieves a strike rate of '*BE* + verb-*ed*' ranging between 50% and 60%. However, if usage analogous to that of the *acclaim* examples listed above is also construed as passive, then in both corpora the percentage of passive structures of *couple* shoots up from 10% to a figure approaching 70%, which would more than suffice to justify the descriptor 'usually passive.' Examples of these in the *British Web 2007* include:

(5) Add to this the plea that the hardship already endured by the people of Britain from the war and this, **coupled** with the continuation of rationing, could make an argument that the benefits of victory were hard to locate.

(6) **Coupled** with our technical skills and expertise, it puts South West Communications at the forefront of technology, with the ability to deliver real value-add services to customers.

(7) The abolition of grants, **coupled** with the introduction of tuition fees, appear to be deterring students, particularly those from poorer backgrounds.

At this point it does not seem unwarranted to imagine that lexicographers' interpretations of passive structures are broader than those of grammarians. Assistance in this regard is not to be found in lexicographers' annotations, as a rule at the beginning of the book, but since one is dealing with a dictionary, it would seem reasonable to look up the meaning of 'passive' in the dictionaries themselves for enlightenment (see Sinclair 1987, 112-113). Here are some definitions:

> *OAL*: (grammar) connected with the form of a verb used when the subject is affected by the action of the verb, for example 'He was bitten by a dog.'

This definition is not very helpful because it fails to exclude reflexives such as *I've cut myself shaving* or *he stabbed himself*, where the subject is incontestably affected by the action of the verb. Further, the *OAL* definition does not embrace impersonal usage such as "It is said the location had been the site of a residence owned by the Kings of Scotland," unless one produces the faintly desperate argument that *it* is affected by *said*.

> *Macmillan*: in passive verbs or clauses, the subject is the person or thing that is affected by the action of the verb. For example in the sentence 'The order was delivered the next day,' the verb group 'was delivered' is passive, and the clause is in the passive voice.

Again, this does not totally exclude reflexive structures or impersonal structures of the type *it is reported that.*

> *Cobuild*: In grammar, the passive or the passive voice is formed using 'be' and the past participle of a verb. The subject of a passive clause does not perform the action expressed by the verb but is affected by it. For example, in 'He's been murdered,' the verb is in the passive.

This is much more specific. The definition excludes reflexives on two counts: (i) passives must consist of '*BE* + past participle,' and (ii) the subject does not perform the action expressed by the verb.

> *CLD*: A passive verb or sentence is one in which the subject does not do or cause the action but is affected by it. For example 'He was released from prison' is a passive sentence.

Again, this excludes reflexives but is less specific than the *Cobuild* definition. It is also worth noting that all the above definitions save that of the *Cobuild*—which specifies that the passive entails a form of *BE*—appear to encompass structures with *get*, *become, seem, look* and *appear* such as the following:

(8) Again, I **got stung** by a jellyfish, but by this time I couldn't have cared less.

(9) These may subsequently **become infected** by common bacteria such as Staphylococcus aureus.

(10) The groups **seemed energized** by my feedback and were keen to complete the questionnaire.

(11) At the final whistle the game could be regarded as pleasant post-season wind-down with South never **looking troubled** by the lower division team.

(12) I tried to **appear fascinated** by this conversation, but admit to wanting him to shut up.

The situation is thus unclear. Aside from the fact that the definitions are all either subtly or radically different from each other, only the *Cobuild* definition expressly bars entry to any structure which is not '*BE* + verb-*ed*,' while the others would seem to embrace a broader spectrum of structures but are accompanied exclusively by examples consisting of '*BE* + verb-*ed*.' This means that despite the discrepancies in descriptions of the passive furnished by both grammars and dictionaries, it is exclusively '*BE* + verb-*ed*' examples which are adopted to

illustrate them. So, are "Acclaimed by so many as the ultimate villa holiday destination, the Algarve is…" and "I got stung by a jellyfish" to be considered passive structures or not? There appears to be no definite answer. And if there is no definite answer, the dictionary label 'usually/often passive' risks being so approximate as to be almost useless.

2.1.2 Passive meaning

To conclude this section on passives, we also need to take into account verb usage which *does* fit the canonical passive construction, but which has what might be termed weak passive meaning, i.e., little correspondence with dictionary definitions of 'passive' which emphasise that the grammatical subject is affected by the action of the verb. Let us take as an example the verb *entitle*. As shown in Table 1 above, the *OAL, CLD, Cobuild* and *Macmillan* all flag this verb as 'often passive' or 'usually passive.' The *Macmillan* entry is as follows:

> **entitle**
> 1.OFTEN PASSIVE to give someone the right to do something
> entitle someone to something: *Membership entitles you to reduced season tickets*.
> entitle someone to do something: *The people who are entitled to vote should be aware of that fact*.
> 2.USUALLY PASSIVE to give a title to a book, poem, or piece of music
> *Her first novel was entitled More Innocent Times*.

Firstly, it is worth pointing out that the presence of the label 'usually passive' accompanying definition 2 reinforces the observations made above with reference to *acclaim* and *couple*. In a 200-line sample concordance of the lemma query *entitle*, 72 occurrences correspond to the meaning described in definition 2. Of these, 68 have passive meaning, but only 3 of these 68 are characterised by the canonical '*BE* + verb-ed,' e.g., "The Marylebone Theatre production **is entitled** 'Rob Roy or The rale Dougal crature' (sic)." If the 200-line sample is representative, then only 3 out of 72 of occurrences of *entitle* correspond—again with reference to sense n.2 above–to the canonical form of the passive and thus to the example provided by the *Macmillan* ("Her first novel was entitled More Innocent Times").

With reference to definition 1 of *entitle*, many occurrences of '*BE* + *entitled'* with the meaning described appear to be characterised by weak passive meaning, for example in the sequence '*BE entitled to feel*.' Here are three of the 46 occurrences of this sequence in the *British Web 2007*:

(13) I think we're **entitled to feel** a little smug when New York steals one of our best ideas.

(14) When you find the captain in his cabin, lying on a bed, with a blanket over his head, and when he says 'wake me up when this is all over,' then you're **entitled to feel** a little nervous.

(15) Maybe the way they act now has something to do with the fact that for three thousand years they were enslaved by trolls and ogres and traded across the globe as commodities? They're **entitled to feel** stroppy!

In these instances the canonical passive structure is present, but the only implicit passive agent imaginable is something vague such as 'the situation,' i.e., "They're entitled [by the situation] to feel stroppy!," and in any case it becomes arduous to find correspondence with the *Macmillan* definition of the passive when it claims that "the subject is the person or thing that is affected by the action of the verb." Of course one might wish to claim this but (i) it seems singularly forced, and (ii) it would open the floodgates, because then all sorts of usage not normally qualified as passive would have to be allowed into the passive club, such as:

(16) **Being depressed** is like any other illness, so you do not need to…

(17) The girls **were** naturally **delighted** at this generous gesture.

(18) The main access to the hall **is stepped**.

In short, *entitle* with the meaning in definition 1 tends to respect the canonical passive construction but in the great majority of instances appears to have negligible passive meaning, while *entitle* with the meaning in definition 2 seldom respects the canonical passive construction but has much stronger passive meaning.

2.2 The labels 'often progressive'/'usually progressive'

Similar drawbacks apply to the labels 'often progressive'/'usually progressive' adopted in learners' dictionaries. Are "Just when we thought there would be no bus coming, the last bus to Mangagoy arrived from Davao," and "The number of people showing up for the so-called parallel summits, i.e., social movements, has seen an even stronger explosion" (with their implicit relative clauses *no bus which was coming* and *people who have been showing up*) to be construed as progressive forms? Further, are elliptical forms to be included? In the *OAL* the verb *kid* is assigned the label 'usually used in the progressive tenses,' but in the *British Web 2007* only around 30% of the occurrences of the verb lemma *kid* comprise the prototypical progressive '*BE* + verb-*ing*' (though there is some interference from *kid* as a noun), for example:

(19) Who do these people think **they're kidding**?

(20) I also know that you will realize that **I'm just kidding**.

Indeed the vast majority of the 132 *-ing* occurrences in the 200-line sample appear in other constructions such as gerunds or elliptical forms:

(21) The teacher's writing on the blackboard. **No kidding**, Anne-Marie sees him.

(22) Love you too, British cousin (**not kidding** either).

(23) And who invented hagis and the bagpipes (**just kidding** scottland is cool). [Original spelling]

Assuming the sample is representative, then 70% of the concordance lines of the verb lemma *kid* do not correspond to the prototypical '*BE* + verb-*ing*.' This outcome must be related in part to the fact that the *British Web 2007* is a corpus of computer-mediated language, and is therefore likely to retrieve more informal use and as a consequence more elliptical structures than other corpora. Nevertheless, one imagines that the compilers of the *OAL*, in order to justify the labelling of *kid* as 'usually used in the progressive tenses,' must be construing 'progressive' very broadly, but this would again be in direct conflict with the *OAL* definition (n.3) of the entry 'progressive':

> **progressive** (also *continuous*) (grammar) connected with the form of a verb (for example *I am waiting* or *It is raining*) that is made from a part of *be* and the present participle.

Although it is perhaps doubtful that dictionary users would actually check such definitions, it goes without saying that learners may be seriously misled by labels referring to passive and progressive, the potential outcome being poor or unusual usage such as "the investigation was coupled by a report" [my example], or a lack of awareness concerning the prolific recurrence of elliptical structures such as "Just kidding."

3. A conflict of form and function

According to Rundell (2008b, 234):

> Though considerable variation still exists among the different MLDs [monolingual learners' dictionaries], what most current coding systems have in common is that they assume very

> little grammatical knowledge on the part of users, and they aim to satisfy users' needs in this department without requiring them to consult explanatory tables and charts.

It is incontestable that the transmission of information in learners' dictionaries is a lot more transparent than it used to be, but when it comes to some of the traditional verb categories adopted, the idea that lexicographers "assume very little grammatical knowledge on the part of users" is hard to uphold. On the contrary, in view of the fact that no clear guidelines are provided, the assumption is apparently that users are rather well-versed in the meaning and scope of grammatical categories. Yet the meaning of the labels 'often/usually passive' and 'often/usually progressive' is opaque, mainly because what falls within the purview of passive and progressive is not established. The absence of information in this regard in turn means that the adverbs *often* and *usually* are not helpful for the user either. See Biber (2012) for further observations.

The difficulty discussed here derives from the precarious relationship between form and function in corpus-based lexicography. The great strength of corpus analysis is the quantification of form, which is performed by the software, not the investigation of function, which is performed by the analyst. As Jones and Waller point out: "Some of the ways we can use the [corpus] data are objective (nobody can dispute the number of times a pattern occurs), but there will always be an element of subjectivity in how we describe and interpret the data" (2015, 16).

The software can retrieve fairly easily, for example all occurrences of '*BE* + verb-*ed*' or '*BE* + verb-*ing*' (form), but the retrieving of passives, progressives etc. (function) depends on analysts, and so is interpretable. For proof of this one need look no further than Table 1 above: lexicographers interpret corpus data in very different ways. In any case, conflict is bound to arise when one adopts traditional grammatical categories to classify the multifarious wealth of data offered by modern corpora. As Sinclair (2004, 173) observes:

> Professional linguists should not be surprised to experience a rather disturbing effect from the massive surge in the availability of evidence and the growing sophistication of the tools for examining it and testing hypotheses against it that corpus linguistics has brought. Some of the vague but useful categories of traditional language analysis, which have served humans well for centuries, are not easily replicated in computational routines; for example 'parts of speech' or 'word class' labelling. Human beings have little difficulty assigning words to a dozen or so word classes, but machines have exposed just how untidy a categorization this is. For English, which has had a lot of attention over many years, there is little or no consensus about how many labels there are […] or how they are defined.

It is thus up to lexicographers to outline with examples, as precisely as possible, what type of usage falls within the compass of passives and progressives. In the absence of such clarification it might be advisable not to include these labels at all.

4. Other verb forms eligible for labelling

Aside from the passive and the progressive, in learners' dictionaries there is scarcely any explicit labelling—again with reference to verbs—of tense, aspect, person, voice or mood. At most we find a sub-entry of the usage in question. What follows is a brief list of verb categories which would appear to deserve labelling at least as much as passives and progressives do. Only the first of these—the imperative—is ever included as a label in dictionaries.

4.1 The imperative

Labels regarding the imperative are rarely adopted in learners' dictionaries. The verb *right-click/right click*, for example, shows a conspicuously high strike rate of over 80% for imperative forms in the *British Web 2007* but is not labelled as often or usually imperative in any dictionary, and there is no sub-entry devoted to its imperative form. The same goes for the verb *click*, which occurs in the imperative almost 60% of the time. As for *hesitate*, which also shows an anomalous recurrence of imperative forms (56%, almost always *don't hesitate*), it is assigned sub-entries for *don't hesitate to do* in *Macmillan*, *LDCE* and *Cobuild*, and a sub-entry for *not hesitate to do* in *CLD*, whereas for the verb *worry*, which has 34% imperative (almost always *don't worry*), only *LDCE* lists the sub-entry *don't worry*. Even *imagine*, at 22% imperative (mostly affirmative), is worth mentioning if one considers that most verbs have a strike rate of something like 5%, but none of the dictionaries assigns it a label or a sub-entry, even though each of them lists at least one example of *imagine* in the imperative. Therefore, the user will occasionally find a sub-entry devoted to an imperative form, but will hardly ever come across an explicit label.

Having said that, the *Macmillan* and *Cobuild* assign the label 'only imperative' or 'usually imperative' to some definitions of *let*, a verb which overall shows a recurrence of more than 30% for imperative, although it increases to over 50% if all occurrences of '*let's* + base form of verb'/'*let us* + base form of verb' are included (*let's go, let us pray*). Interestingly, of the 10 definitions of *let* included in the *OAL* (I exclude the sense corresponding to 'rent out'), 8 of them bear the label 'no passive' and yet, although around 90% of the examples supplied in the *OAL* for the various definitions of *let* are imperative, the label 'often imperative' is absent. It seems paradoxical that a feature which is consistently absent should be labelled (as absent) eight times within a single

verb entry, while a feature which is consistently present—at least in the examples listed in the *OAL*—is not labelled at all.

4.2 The simple present

The *Longman Grammar of Spoken and Written English* (1999, 459) reports that the simple present tense "is strongly associated with verbs denoting mental and logical states" particularly those expressing emotions and attitudes, and the verbs listed include *bet, doubt, care, fancy, reckon, suppose*; see also Biber (2012, 10). Learners' dictionaries do not explicitly label verbs adopted primarily in the simple present, though often the examples provided would suggest that its strike rate is high for certain verbs. Let us take as an illustration of this the verb *tend*. In the *British Web 2007* the simple present of the verb *tend* accounts for around 80% of all the occurrences of the verb, and it accounts for almost all of the occurrences of *tend* when it is followed by the *to*-infinitive. If we focus on this latter structure:

(24) However, metaphors also **tend to import** unwanted meanings.

(25) Petitions by themselves **don't tend to achieve** much.

we find that in examples provided by learners' dictionaries the simple present practically has a monopoly: of the 31 examples of '*TEND* + *to*-infinitive' furnished across the *OAL*, *LDCE*, *Macmillan*, *Cobuild* and *CLD*, 30 are in the simple present. Despite this, *tend* is not labelled for simple present in any of the dictionaries. Similarly, the verb *need* occurs in the simple present 65% of the time in the *British Web 2007*, and although a substantial number of the examples listed for this verb in the dictionaries are simple present, it is never flagged. Again, this is significant because most verbs have a strike rate of less than 10% for simple present.

It is worth noting in passing that the choice of examples in dictionaries is not a consistently reliable indicator of effective distribution. The examples of the verb *suit* in the *OAL* and *Macmillan* are dominated by the simple present, but in the *British Web 2007* the occurrences in the simple present are only 15%. Indeed, in the corpus it is the infinitive of this verb which dominates, particularly the *to*-infinitive, for instance "there's a range of restaurants to suit all tastes."

Here too one is faced with the problem of defining what verb categories include and exclude, a question discussed in 2.1.1 above with reference to the passive. In this section I am assuming that the simple present embraces negative and interrogative forms, but in the *OAL* the grammatical meaning of the adjective *simple* is:

> used to describe the present or past tense of a verb that is formed without using an auxiliary verb, as in She loves him (= the simple present tense) or He arrived late (= the simple past tense).

This definition excludes present and past interrogatives (*does she know?, did she win?*), negatives (*she doesn't know, she didn't win*), and emphatic use such as *I do write to her* [my examples], an exclusion which seems counter-productive. If then we look up 'present tense' in the *OAL*, we find that it is defined as "the form of a verb that expresses an action that is happening now or at the time of speaking," and here too there are difficulties. In particular, neither of the two definitions cited appears to exclude the gerund in, for example, "I firmly believe, speaking generally, that the aristocracy will sustain no injury from it whatever," because in this sentence the verb structure does not feature an auxiliary verb and describes an action that is happening at the time of speaking. I shall not refer to definitions of grammatical categories any further, but once again Sinclair's observations cited in section 3 regarding traditional grammatical categories ring very true. However, there is nothing to prevent dictionary compilers providing illustrative examples of what they consider to be passive, simple present etc. in the key at the front of the dictionary.

4.3 The simple past

It comes as no surprise that the English simple past tense is particularly recurrent in fiction (see the *Longman Grammar of Spoken and Written English*, 459) and, as a consequence verbs such as *exclaim* and *retort* have high percentages of simple past tense forms. The simple past accounts for 78% of the verb lemma *exclaim* in the *British Web 2007* (e.g., "'I don't believe him!' exclaimed Miss Pole, in a defiant manner"), and of the 13 examples of this verb listed across the *OAL*, *LDCE*, *Macmillan*, *Cobuild* and *CLD*, 11 are in the simple past. A similar situation applies to the verb *retort*: 65% of simple past ("'You entered my kingdom uninvited,' retorted the king, 'and you tell me to my face I am ugly'") and in the five learners' dictionaries referred to above, there are 12 examples in the simple past out of a total of 13. Once again, a single tense dominates the distribution—most verbs have a frequency rate of between 10% and 20% for simple past–and yet there is no label to signpost this in dictionaries.

4.4 Perfective forms

The verb *pledge* occurs in perfective forms at an abnormally high rate of 30%, for example:

(26) A sizeable bond **has been pledged** to safeguard our clients.

(27) The Chamber **has pledged** to work with the SEC.

(28) Something the government **had pledged** not to do in its Election Manifesto.

The majority of verbs have a strike rate for perfective forms of around 4-5%, so *pledge* is certainly anomalous and would merit a label in this regard. The verb *elude* (26%) also shows a high percentage:

(29) Billy Childish is the recipient of an international cult status that seems (so far) to **have eluded** him in the UK.

(30) Although the ultimate prize **has eluded** generations of Scottish golfers down the years, there have been many fine exponents of the game.

Other verbs with notable percentages of perfective forms include *campaign* (21%), *implicate* (21%), *appoint* (19%) and *master* (18%), but once again dictionaries provide no explicit indications. By contrast, other verbs tend to steer clear of perfectives, for instance *deserve, depend, hope* and *contain*, which have extremely low percentages.

4.5 Grammatical person

Having carried out corpus analyses, albeit cursory, of verb forms relating to voice, aspect, tense and mood, it would seem natural to continue the investigation in terms of grammatical person. This task is, however, arduous, because aside from the present third-person singular, English does not possess distinct endings for each grammatical person, with the consequence that for any given verb the corpus analyst is required to trawl all forms of the verb in the concordance aside from the *to*-infinitive. With this in mind it is perhaps no surprise that not even the *Longman Grammar of Spoken and Written English*, normally so assiduous in providing frequency percentages of words within grammatical categories, supplies statistical data concerning verb person and number.

4.5.1 The first person

Predictably, first-person forms are a recurrent feature of verbs expressing opinions and wishes, and this is particularly true of the first-person singular. Some examples are presented in Table 2.

	% of first-person singular	% of first-person plural
hope	38	23
guess	58	2
doubt	45	6
believe	30	17
reckon	29	8
wish	12	2
prefer	19	3
confess	14	1

Tab. 2: Examples of verbs expressing opinions and wishes: percentages of first-person forms in the *British Web 2007* corpus

Table 2 contains some significant figures: my corpus searches would suggest that most verbs show a percentage of 2-3% of singular/plural first-person forms, but for instance both *hope* and *guess* achieve a percentage of around 60%. Dictionaries do not provide labels to signpost abundant first-person use, though sub-entries devoted to first-person forms are sometimes present, notably *I guess*, which is awarded a sub-entry in all five of the learners' dictionaries consulted.

4.5.2 The third person

Whereas conspicuous recurrence of second-person (non-imperative) forms is dominated by fixed phrases such as *if you prefer, if you like* and *as you wish*, the analysis of the distribution of third-person forms is greatly affected by impersonal structures (*it would seem that…, it was hoped that…, it's snowing*), but it is worth noting that many verbs are dominated by non-impersonal third-person usage—the verb *tend* once again shows a very high percentage, in this case around 90%. As implied above, investigating this situation with regard to English would require the work of a team rather than of a single researcher so I shall not pursue it here, though of course verbs governed by exclusively or mostly inanimate subjects, from *cost* through to less frequent verbs such as *leak*, *drip* and *jackknife*, show very high figures for third person.

5. That which is not: the labels 'no passive' and 'no progressive'

In section 4.1 above on the imperative, I considered the entry for the verb *let* in the *OAL*, remarking upon the repeated use of the label 'no passive.' On the whole, this strategy of

highlighting features which do *not* characterise a given verb is not frequent in learners' dictionaries. Most of them state, for example, that *deem* and *need* are not used in the progressive, but let us consider the question in more detail for verb *resemble* in Table 3.

OAL	No passive	No progressive
Macmillan	No passive	-
Cobuild	-	No progressive
CLD	-	-
LDCE	-	No progressive

Tab. 3: Labels 'no passive' and 'no progressive' assigned to *resemble* in learners' dictionaries

Again the outcomes in Table 3 must be the result of differing methods and/or parameters across dictionaries, but perhaps more striking is the fact that for the verb *elude*—of which I can identify only a handful of passive occurrences in the entire *British Web 2007* even when adopting the broadest of interpretations of 'passive'—only the *Cobuild* signals it as not passive. Naturally a degree of inconsistency is to be expected: identifying what something is not rather than what it is takes us into massive unchartered territory, but curiously it is once again the passive and progressive which are singled out for attention (in this case even by the *LDCE*, which as pointed out in Section 2, is parsimonious in its use of labels such as these). Table 3 suggests that *resemble* is used very sparingly in the passive and progressive, but on closer inspection one discovers that it is also used very sparingly in the imperative, in perfectives, in constructions indicating future time and in the first person, none of which gets a mention in dictionaries. The fairly high-frequency verb *reckon*—18,392 hits at 11.88 words per million in the *British Web 2007*–is scarcely ever used in structures denoting future time (*will/shall reckon*, *BE going to reckon* or even *BE about to reckon*), and *allege* has comparatively low frequency rates for the imperative, the infinitive, forms of the future, the conditional, the first person and the second person, but these are not labelled either. I am certainly not suggesting that they should be, but the discrepancy between the lexicographical treatment of the passive and progressive by comparison with other verb categories is striking.

6. Limitations of the research

A survey as brief as the one offered in this paper is bound to have shortcomings. The principal limitations are the following:

(i) Ideally, any discussion of word frequency within grammatical categories should factor in criteria such as register, spoken vs written, formal vs informal (see Biber 2012; Jones and Waller 2015, 121-127). Having said that, the grammar labels examined in the five dictionaries mostly apply to headwords and not to sub-entries, and are therefore rather general by nature.
(ii) Linked to the previous point, I have not as a rule considered how frequency rates of words in grammatical categories are contingent upon the different meanings of the individual dictionary headwords. As stressed by the lexicographers Atkins and Rundell (2008, 399):

> As we convert the information in the database into a final entry, it's important to distinguish between features which apply to the entry as a whole, and features that belong to specific senses.

This applies, for example, (a) to the verb *slate*, which occurs more often in canonical passive structures when its meaning is synonymous with *plan* ("a release has been promised for many months, and is now slated for August") or with *suggest somebody for a job or position* ("Deputy Defense Minister Paul Wolfowitz, who is slated to become the next head of the World Bank"), than when its meaning corresponds to *criticise*, since in this sense there is a much higher percentage of active forms ("Manchester United defender Mikael Silvestre has slated the current Arsenal team as the worst he's faced"), and (b) to the verb *bet*, which shows an abnormally high percentage of simple present first person singular when the sense is *be certain* ("I've never seen them live myself, but I bet they are bloody brilliant") by comparison with its meaning of *gamble* ("I bet you five pounds it will turn out as I say").
(iii) The statistical data supplied in this paper derive almost exclusively from a single general web corpus (the *British Web 2007)*, but of course a different corpus would have generated different outcomes. For example, in the *British National Corpus* the verb *hesitate* has a strike rate of 70% for the past simple tense, whereas the *British Web 2007* shows just 13%, and this is because the *British National Corpus* contains proportionally so much more narrative than the *British Web 2007* – a typical use of *hesitate* in narrative is immediately before inverted commas: "He hesitated, 'We're fools in love, the pair of us" (*British National Corpus*). And this will in turn affect the percentages of the other forms of *hesitate*: in the *British National Corpus* the strike rate of the imperative (mostly *don't hesitate*) is 9%, whereas, as noted in 4.1 above, the corresponding rate in the *British Web 2007* is 56%. Of course, it is the use of different corpora on the part of the compilers of the various dictionaries discussed that helps to explain why the respective findings are occasionally so diverse.

(iv) The percentages assigned to verb forms within grammatical categories are based on 200-line random samples from the corpus, a methodology which has certain drawbacks. Firstly, some words have much higher overall frequency than others, for example a 200-line sample of *procrastinate*, which has a total of 453 occurrences in the *British Web 2007*, is likely to be more representative of the total compared to a 200-line sample of *receive*, which has a total of over half a million, and this may impinge upon the statistical validity of the findings. Secondly, although in Sketch Engine a random sample of 200 will always produce the same outcome providing that the corpus itself has not been not altered, a random sample of, say, 199 or 201 will depart to some degree from the sample of 200 and thus produce different results. Again, the different sampling methods adopted by the compilers of learners' dictionaries may in part account for the recurrent discrepancy of the outcomes. Thirdly, the tagging of the *British Web 2007* shows some flaws which, if undetected, may radically affect findings. For example, a lemma query of the verb *prize* (a verb considered to be usually passive by some learners' dictionaries—see Table 1 above) captures a large number of examples of the noun, particularly the plural noun *prizes*. If these are not carefully filtered out then, for instance, the percentage of passive instances calculated will be severely skewed.

(v) I have included some verb categories in the analysis and excluded others. For reasons of time and space, the infinitive, the conditional, and forms referring future time have for instance not been examined.

7. Conclusions

Monolingual learners' dictionaries constitute a pioneering step forward for English linguistics, and the inclusion of detailed but clearly-presented grammatical information has been one of their most important contributions. Grammar labels for dictionary headwords can relate primarily (i) to the structures that accompany the specific headword, notably the labels transitive/intransitive, or (ii) to the form of the headword itself, for instance the labels uncountable and imperative. With specific reference to verbs, it is striking that in learners' dictionaries the great majority of labels of this second type relate to the (abundance or scarcity of) passive and progressive forms, yet other verb categories would appear to be equally eligible for labelling in this way, for instance the simple present, perfective forms and the first person, since some verbs show conspicuous percentages for these categories.

The manifest prioritising of passive and progressive constructions is curious, particularly as there are legitimate doubts as to what these two labels really mean in dictionaries and therefore what structures they encompass. Such doubts emerge at least in part from a sometimes

precarious relationship between form and function in corpus-based lexicography and indeed in corpus investigations *tout court*.

While one accepts that from one scholar to the next there are bound to be different interpretations regarding the breadth of categories such as passive and progressive, it would seem a simple enough task for lexicographers to define and exemplify the labels adopted in the key accompanying the dictionary, particularly as definitions of grammatical categories within the dictionary itself are not always illuminating.

Finally, outcomes regarding the frequency of passives, progressives and other word classes are contingent not only upon what usage these encompass, but also upon factors such as the specific corpora consulted, and the sampling methods adopted. All this has significant repercussions for the information consigned to learners and consequently for their language performance.

Dominic Stewart *teaches English Language and Linguistics and Italian-English Translation at the University of Trento. He has published primarily in the field of corpus linguistics and translation. His publications include* Semantic Prosody: A Critical Analysis *(2010) and* Italian to English Translation with Sketch Engine: A Guide to the Translation of Tourist Texts *(2018).*

Works cited

Atkins, Sue and Michael Rundell. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press, 2008.

Biber, Douglas. "Register as a Predictor of Linguistic Variation." *Corpus Linguistics and Linguistic Theory* 3.2 (2012): 9-37.

Biber, Douglas, et al. *Longman Grammar of Spoken and Written English*. London: Longman, 1999.

Bogaards, Paul and Willem A. Van der Kloot. "The Use of Grammatical Information in Learners' Dictionaries." *International Journal of Lexicography* 14.2 (2001): 97-121.

Cambridge Learner's Dictionary Online. https://dictionary.cambridge.org/it/dizionario/learner-inglese. Last visited 02/07/20.

Collins Cobuild Reverso Dictionary. https://dictionary.reverso.net/english-cobuild. Last visited 02/07/20.

Dziemianko, Anna. *User-friendliness of Verb Syntax in Pedagogical Dictionaries of English*. Tübingen: De Gruyter, 2006.

Hands, Penny. *Collins Cobuild English Grammar*. Glasgow: Harper Collins, 2017.

Hanks, Patrick. "The Corpus Revolution in Lexicography." *International Journal of Lexicography* 25.4 (2012): 398-436.

Jackson, Howard. *Lexicography: An Introduction*. London: Routledge, 2002.

Jones, Christian and Daniel Waller. *Corpus Linguistics for Grammar: a Guide for Research.* London: Routledge, 2015.

Kilgarriff, Adam. "Putting Frequencies in the Dictionary." *International Journal of Lexicography* 10.2 (1997): 135-155.

Longman Dictionary of Contemporary English Online. https://www.ldoceonline.com. Last visited 02/07/20.

Macmillan Dictionary Online. https://www.macmillandictionary.com. Last visited 02/07/20.

Oxford Advanced Learner's Dictionary Online. https://www.oxfordlearnersdictionaries.com/?cc=it. Last visited 02/07/20.

Rundell, Michael. "More Than One Way to Skin a Cat: Why Full-Sentence Definitions Have Not Been Universally Adopted." *Practical Lexicography: A Reader*. Edited by Thierry Fontenelle. Oxford: Oxford University Press, 2008a. 197-209.

---. "Recent Trends in English Pedagogical Lexicography." *Practical Lexicography: A Reader*. Edited by Thierry Fontenelle. Oxford: Oxford University Press, 2008b. 221-243.

Sinclair, John M. "Grammar in the dictionary." *Looking Up: An Account of the COBUILD Project in Lexical Computing*. London: Collins ELT, 1987. 104-115.

---. *Trust the Text: Language, Corpus and Discourse*. London: Routledge, 2004.

Swan, Michael. *Practical English Usage*. Oxford: Oxford University Press, 2017.