

# The Inclusion of Word Formation in OALD8: The Case of Undefined Run-ons

Alenka Vrbinc (*alenka.vrbinc@ef.uni-lj.si*), Faculty of Economics,  
and  
Marjeta Vrbinc (*marjeta.vrbinc@ff.uni-lj.si*), Faculty of Arts  
*University of Ljubljana, Ljubljana, Slovenia*

---

**Abstract:** The study presented in this contribution aims to investigate whether the print edition of OALD8 still sticks to the tradition of including many derivatives as run-on entries. For the purpose of the study, a database was compiled consisting of 1,200 lemmata with full entry status, and in this lemma range, 145 run-ons were found. The parts of speech of the lemma under which a run-on could be found were closely studied, and so were the illustrative examples. Nouns are most commonly included as run-ons, followed by adverbs, adjectives and verbs. The problem of polysemous entries at the end of which undefined run-ons can be found is also discussed. When it is difficult to draw parallels between the sense(s) of the lemma and the run-on deriving from it, it would be much better to include the run-on as the lemma and to define all its senses separately or to include undefined run-ons at the end of each individual sense of the lemma. Finally, it should be clear to the user when and why a certain word is included as a run-on and not as a lemma, and how s/he can establish a relation between the meaning of the lemma and the run-on.

**Keywords:** MONOLINGUAL LEARNER'S DICTIONARIES, OALD8, LEMMATA, SECONDARY LEMMATA, UNDEFINED RUN-ONS, PROPORTION OF RUN-ONS TO ENTRIES, EXAMPLES OF USE, DERIVATIVES, SUFFIXES, PART-OF-SPEECH LABELS, GRAMMATICAL INFORMATION

**Opsomming: Die insluiting van woordvorming in die OALD8: Die geval van onverklaardes.** Die studie waaroor daar in hierdie artikel berig word, probeer vasstel of die gedrukte weergawe van die OALD8 die tradisie handhaaf om baie afleidings as onverklaarde lemmata op te neem. Vir die doel van hierdie studie is 'n databasis saamgestel van 1,200 lemmata met volle artikels en tussen hierdie lemmata is 145 onverklaarde lemmata gevind. Die woordsoort van die lemmata waaraan onverklaardes geheg is, is noukeurig nagegaan asook die voorbeeldmateriaal. Naamwoorde tree die algemeenste as onverklaardes op, gevolg deur bywoorde, adjektiewe en werkwoorde. Die probleem van polisemiese inskrywings waaraan onverklaardes geheg word, word ook bespreek. Wanneer dit moeilik is om die verband tussen die betekenisonderskeidinge van 'n lemma en die voortspruitende onverklaardes te lê, sal dit beter wees om die onverklaardes as volle lemmata op te neem en die onderskeie betekenisonderskeidinge te verklaar, of om die onverklaardes aan die einde van elke verklaring van 'n betekenisonderskeiding te heg. Dit moet vir die gebruiker duidelik wees waar en wanneer 'n woord as onverklaarde opgeneem word en nie as volle lemma nie en ook hoe die verband tussen die betekenis van die lemma en die onverklaarde vasgestel kan word.

**Slutelwoorde:** EENTALIGE AANLEERDERWOORDEBOEK, GEBRUIKSVORBEELDE, GRAMMATIESE INLIGTING, LEMMATA, OALD8, ONVERKLAARDES, SEKONDÊRE LEMMATA, SUFFIKSE, VERHOUDING VAN ONVERKLAARDES TOT INSKRYWINGS, WOORDSOORTETIKETTE

## 1. Introduction

In the A–Z part of any dictionary, users can find lemmata listed in alphabetical order and accompanied by many pieces of information. Dictionary users definitely expect that a dictionary entry will present the facts as clearly as possible. The amount of information included in a particular dictionary depends a great deal on the type of dictionary; moreover, the needs and skills of potential target users based on the user profile should also be taken into account (Atkins and Rundell 2008: 200). Information about a lemma is especially detailed in monolingual learner's dictionaries, which are intended not only for decoding but also for encoding purposes. Learners of a foreign language need in-depth information that enables them to use a word correctly in context, which means that besides the meaning, they are also interested in pronunciation, grammatical information, grammatical patterns, collocations, synonyms, usage, etc. Apart from lemmata, two other types of lexical items within the entry should be mentioned: secondary lemmata and run-ons. What these have in common is that they both appear at the end of the entry in a special section that is often marked by a special symbol. However, the treatment of secondary lemmata differs from that of run-ons. Secondary lemmata, e.g., idioms and phrasal verbs, are defined and mostly also exemplified, whereas run-ons are not explained. What users learn about them is their part of speech; grammatical information is also provided, and sometimes their use is illustrated by means of one or more examples. Since the aim of this contribution is to study the inclusion and treatment of run-ons in OALD8, we will now take a look at some basic issues regarding run-ons.

Studying the existing literature shows that lexicographers are aware of a problem with the inclusion of run-ons. Stein (2002: 62), for example, points out that "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". Similarly, Béjoint (2010: 283) notices that "some items in the main A–Z list are treated as if they were of lesser importance, because they can easily be accessed via other words". Atkins and Rundell (2008: 237), on the other hand, emphasize the fact that run-ons can only be used in monolingual dictionaries when the word form is infrequent; when its meaning is unambiguously deducible through the application of basic word-formation rules; when its pronunciation can be predicted from the pronunciation of the headword to which it is attached; and when its grammatical and collocational behaviour is simple and predictable. Stein (2002: 50) also agrees that criteria should be determined

in order to decide under which conditions a certain derivative will be treated as a lemma and under which as a run-on entry. She mentions lexicalization as one of the criteria, but at the same time, she observes many inconsistencies when comparing existing dictionaries.

The most important question that should be posed is whether dictionary users can benefit at all from run-on entries. Stein (2002: 11) believes that this practice can somehow be justified in dictionaries intended for foreign learners because foreign learners are told that certain coinages exist in the language. On the other hand, she claims that native speakers would not really care whether or not derivatives ending in a specific suffix exist. Svensén (2009: 376) agrees with Stein in that nesting has a pedagogical purpose because bringing together lemmata that are morphologically related can promote word comprehension and vocabulary learning. It is, however, doubtful whether foreign learners could apply theoretical knowledge about word formation when looking up undefined run-ons. That is why Atkins and Rundell claim (2008: 235) that "run-on entries tend not to be used so much in learner's dictionaries, the idea being that learners have enough trouble finding what they want without having to burrow around in an entry of a headword that is not the object of their search". Also Stein (2002: 12) notices that some lexicographers and publishers are already trying to get away from run-on practice, the reason being that it is not user-friendly. According to Bergenholtz and Gouws (2013: 72), the use of a cumulative list of unexplained word formations with the lemma sign can be useful for the native speaker who only wants to check whether the word formation exists, how it is spelt or whether it is used with a linking morpheme or not.

Landau (2001: 102) believes that many dictionaries include even derivatives that may never have been used, simply to increase the number of entries in the dictionary. The selection of derivatives listed as run-ons is often arbitrary, thus making consultation of run-on entries difficult (Hartmann and James 1998: 121). Landau (2001: 102) points out that there are cases when the run-ons are more commonly used than the words to which they are run on, and the same observation is made by Atkins and Rundell (2008: 237). This means that lexicographers do not always regard the frequency of a certain word as the main criterion for inclusion of either the lemma or the run-on. Large linguistic corpora now enable lexicographers to determine the relative frequency of a particular word or of a particular sense, but according to Landau (2001: 104), relatively high frequency does not always guarantee lemma status, although it is obviously of great importance and is one of the most important factors to be considered. Atkins and Rundell (2008: 237) also draw attention to the fact that some derivatives are simply used too frequently to be handled as run-ons, and they may also show signs of unpredictable behaviour. Another problem the lexicographers encounter when including a certain word as a run-on is that a run-on may have acquired senses not adequately covered by the root word (Landau 2001: 102).

Traditionally, the main reason some words are included as run-ons is to conserve storage space, which is relatively restricted in traditional paper dic-

tionaries, as opposed to electronic dictionaries where it is relatively unrestricted, thanks to modern, high-capacity storage media (Lew 2011: 4). That is why the aim of our study is:

- to investigate whether the print edition of OALD8 still sticks to the tradition of including many derivatives as run-on entries,
- to gain insight into the proportion of run-ons to entries,
- to research what words merit the status of an undefined run-on and to which lemma the derivative is run on and, last but not least,
- to study additional information found in the run-on entries.

## 2. Methodology

In order to study the run-ons in OALD8, a database was compiled consisting of 1,200 lemmata, each with the status of a full entry. Six hundred lemmata were taken from a randomly chosen stretch in the letter C (i.e., from *caterer* to *chicken wire*) and six hundred from a randomly chosen stretch in the letter S (i.e., from *shwa* to *slip knot*). The material for the study was chosen on the basis of the following criterion: the letters C and S belong to two letters in English with the greatest number of words, thus according to Bukowska (2010: 1265) exhibiting greater variation than letters that have lower numbers of words. Apart from that, her recommendation is to choose a simple random selection of pages (ibid: 1267). Phraseological units (included in the idioms section) and multi-word verbs (included in the phrasal verbs section) that have the status of secondary lemmata were not included in the database, the only exception being those lemmata marked as verbs used in combination with particles only, which means that they appear only as a multi-word verb (e.g., *sick* is labelled as an adjective, noun and verb, but under the verb, only the multi-word verb *sick sth up* is entered and treated). All lemmata were typed into the database together with a part-of-speech label. If the entry contained an undefined run-on, the latter was also entered into the database together with a part-of-speech label. All run-ons in the database were then counted in order to determine the proportion of run-ons to entries. The total number of undefined run-ons in our database amounts to 145. The next step was to closely study the types of run-ons as regards their part of speech as well as the part of speech of the lemma under which a run-on could be found. Since the use of many undefined run-ons is illustrated by examples, these examples were also included in our database and were studied more closely to see whether they were sufficiently informative and intelligible to enable the user to better understand and use the word in question.

## 3. Results

The six hundred lemmata beginning with the letter C include 58 undefined

run-ons. The run-on *Central American* is counted twice, since it contains a multiple PoS label, i.e., adjective, noun. In the case of the 600 lemmata beginning with S, they include 87 undefined run-ons. Two run-ons, i.e., *sixteenth* and *sixtieth*, have multiple PoS labels, i.e., ordinal number, noun, and are therefore counted twice. If we compare the two randomly chosen stretches separately, it can be seen that there are far more undefined run-ons in the stretch with the letter S than in the stretch with C. However, since the database should be treated as a whole, it can be determined that the proportion of undefined run-ons to lemmata is 145 to 1,200, which means 1 : 8.3 and that 12.1% are indeed undefined run-ons. Table 1 shows the number of lemmata, the number of undefined run-ons by individual stretches and as a total and the proportion of lemmata to run-ons.

	Lemma	Undefined run-on	Proportion of lemmata to run-ons
Letter C	600	58	10.3 : 1
Letter S	600	87	6.9 : 1
<b>Total</b>	<b>1,200</b>	<b>145</b>	<b>8.3 : 1</b>

**Table 1:** The number of lemmata, the number of undefined run-ons and the proportion of lemmata to run-ons

Since 12.1% of all lexical items in our database constitute run-ons, it is now necessary to take a look at the parts of speech of the undefined run-ons. It can be seen that nouns are most commonly included as run-ons (i.e., 24 nouns beginning with C, 44 beginning with S, the total being 68 nouns or 46.9%). Nouns are followed by adverbs (i.e., 21 adverbs beginning with C and 26 beginning with S are included as run-ons, the total being 47 adverbs or 32.4%). Adjectives appear less frequently as undefined run-ons (i.e., 10 adjectives beginning with C and 12 beginning with S, the total being 22 adjectives or 15.2%), whereas verbs are rarely found as run-ons (i.e., 3 verbs beginning with each of the two letters, the total being 6 verbs or 4.1%). The only other part of speech that can be found as an undefined run-on in our database is ordinal numeral, which can be found only twice in the letter S (1.4%). Table 2 shows undefined run-ons by part of speech, by each letter and as the total number.

	Undefined run-ons by part of speech				
	Nouns	Adverbs	Adjectives	Verbs	Ordinal numerals
Letter C	24	21	10	3	/
Letter S	44	26	12	3	2
<b>Total</b>	<b>68 (46.9%)</b>	<b>47 (32.4%)</b>	<b>22 (15.2%)</b>	<b>6 (4.1%)</b>	<b>2 (1.4%)</b>

**Table 2:** Undefined run-ons by part of speech

It is necessary to take an in-depth look at the run-ons by parts of speech. As is evident from our database, the most frequent suffix attached to nouns entered as run-ons is *-ness* (20 nouns); this suffix is followed by *-ing* (13 nouns); the suffix *-er* is also quite common (11 nouns), whereas all other suffixes appear rarely (3 nouns ending in *-ity*, 2 nouns in *-th*, and only one noun each can be found ending in *-ization*, *-ability*, *-ment*, *-ist*, *-an*, *-cy* and *-ism*).

Apart from nouns derived by suffixes, our database also contains nominal undefined run-ons that can be regarded as cases of conversion, which means that they are homonymous with the lemma to which they are run on. Twelve nouns of this type were identified, e.g., *centigrade*, *centre-left*, *simper*, *simulcast*.

As regards the part of speech of the lemma to which nouns are run on, it can be seen that exactly half of all undefined nouns (i.e., 34 out of 68 nouns) appear at the end of the entry for an adjective with which they are semantically linked (e.g., *cautiousness* under *cautious*, *silliness* under *silly*); slightly less than a third (i.e., 20 nouns) are listed at the end of the entry for a verb (e.g., *centralization* under *centralize*, *signposting* under *signpost*), and slightly more than a sixth (i.e., 12 nouns) are enumerated at the end of the entry for another noun (e.g., *Catholicism* under *Catholic*, *sightseer* under *sightseeing*). Only two nouns are run on at the end of the entry for a numeral (i.e., *sixteenth* and *sixtieth*, in the entries for *sixteen* and *sixty*, respectively).

As regards the number of adverbial run-ons, they are second only to nouns. With very few exceptions (*centre stage*, *sidelong*, *sideward* (with its variant *sidewards*), *silky* and *sky-high*), the adverbs that are included in OALD8 and appear in our database end in *-ly*. An adverbial run-on that is included at the end of the entry for any other part of speech rather than under an adjective with which it is semantically linked can be regarded as the exception rather than the rule. Only four out of 47 adverbs in our database can be found in entries for other parts of speech: nouns (*centre stage*, *chauvinist*), verbs (*simper*) and other adverbs (*single-handed*). All other adverbs are included at the end of the entry for an adjective that can either be the root word used to form an adverb by adding *-ly* (e.g., *cautiously* is entered under the adjective *cautious*, *shyly* under *shy*, etc.) or a homonym (e.g., *sidelong*, *sky-high*).

As far as adjectives treated as undefined run-ons are concerned, it can be said that they are mostly included in the entries for the nouns to which they are semantically linked (18 out of 22 adjectives, e.g., *centralist* under the noun *centralism*, *slandorous* under *slander*), and very rarely under the verbs (only two adjectives, i.e., *cheering* under the verb *cheer* and *slighting* under the verb *slight*) or adverbs (only two adjectives, i.e., *sideways* and *single-handed* under their homonyms *sideways* and *single-handed*). The suffixes attached to adjectival undefined run-ons differ greatly, and in the majority of the adjectives in our database, only one example can be found per suffix. The suffixes are as follows: *-ic*, *-(e)d*, *-(i)an*, *-ist*, *-ing*, *-y*, *-al*, *-ous*, *-ly*, *-less*.

In our database, verbs rarely appear as undefined run-ons. There are only six verbs treated in this way, i.e., *catnap* (in the entry for the noun *catnap*), *caw*

(in the entry for the noun *caw*), *checkmate* (in the entry for the noun *checkmate*), *sideswipe* (in the entry for the noun *sideswipe*), *sightsee* (in the entry for the noun *sightseeing*) and *silk-screen* (in the entry for the noun *silk screen*). As can be seen, they are included in the entries for the nouns with which they are homonymous, the only exception being the verb *sightsee*.

Another interesting issue is that undefined run-ons are followed by different types of information. Some are only listed at the end of the entry for the lemma and are accompanied by a PoS label and (possibly) by grammatical information, whereas others also include examples illustrating the use of the run-on. There are 26 out of 58 run-ons beginning with the letter C with illustrative examples, and the same holds true of 31 out of 87 undefined run-ons beginning with the letter S.

	Undefined run-ons	Undefined run-ons + illustrative example(s)
<b>Letter C</b>	58	26
<b>Letter S</b>	87	31
<b>Total</b>	145	57

**Table 3:** Number of undefined run-ons and number of undefined run-ons with illustrative examples

#### 4. Discussion

Since many words such as nouns ending in *-ance*, *-ence*, *-ency*, *-er*, *-ing*, *-ness*, *-ity*, *-ism*, *-tion*, etc., adjectives in *-able*, *-al*, *-ary*, *-ful*, etc. and adverbs ending in *-ly* (Svensén 2009: 132; Béjoint 2010: 284) can be defined by certain patterns that repeat (e.g., *-ness* is supposed to mean "the state of being ...", *-ly* is supposed to mean "in a ... manner", etc.) (cf. also Landau 2001: 102), space in print dictionaries is often saved by including such derivatives at the end of the entries for the related lemmata as undefined run-ons. Space saving is no longer an issue in electronic or online dictionaries, but the subject of our investigation is a print dictionary, where space is still a factor that should be taken into account. However, if the undefined run-ons found in the print edition are checked in the online version of OALD as well as the CD-ROM version accompanying the print edition, it can be seen that the same lexical items are included as run-ons as in the print edition, even though there are no space limitations in electronic dictionaries. The only difference is that the online version and the CD-ROM version include illustrative examples in some run-ons which are not exemplified in the print edition (e.g., *shyness* and *shyly*). Dictionary users do not concern themselves with problems of space or other limitations, but they may be faced with difficulties when performing a look-up operation, since they most certainly expect that the word they seek will be included and treated as a lemma and not as a run-on. A lemma is easy to find because it is part of the

dictionary macrostructure, it is precisely defined, grammatical codes and grammatical patterns are provided, the use of the lemma is exemplified, collocations are included, etc. On the other hand, an undefined run-on is much more difficult to find, as it is hidden within the dictionary microstructure, and far fewer pieces of information about it can be obtained. (These considerations apply to print dictionaries only, since finding a lexical item in an electronic dictionary poses no problems at all.) Apart from that, it is implicitly assumed that users are capable of an immediate constituent analysis (Stein 2002: 54) when they come across a derivative listed as a run-on. What may be particularly troublesome is the absence of definitions, since dictionary users are mostly interested in the meaning of a word when they look it up (cf. Atkins and Rundell 2008: 405). Another issue that should be addressed is the criteria for including a certain word either as a lemma or as a run-on. Doubtlessly, dictionary users cannot possibly be expected to know how the lexicographers determine the status of a particular lexical item, which means they do not know where to look up a word — in the macro- or microstructure.

In the print edition of OALD8, the policy of including run-ons is briefly mentioned in the *Key to dictionary entries* (p. ix):

Some words that are **derivatives** of other words do not have their own entry in the dictionary because they can be easily understood from the meaning of the word from which they are derived (the root word). They are given in the same entry as the root word, in a specially marked section.

Users of the CD-ROM and the online versions, however, are completely deprived of this information. Nowadays, in the age of modern technology, it would be over-optimistic to assume that the majority of dictionary users would consult a print edition of any dictionary, since the advantages of electronic versions are so numerous that they far outweigh the benefits of print dictionaries.

As can be seen in the Results section, many nominal run-ons are characterized by typical suffixes. It is, however, questionable whether an average dictionary user is able to grasp the meaning of the derivative, since it does not suffice to know the meaning of the lemma from which the undefined run-on is derived — the user also needs to be familiar with the meaning of the suffix. For example, if one looks up the suffix *-ing* in OALD8, one learns that it is 'used to make the present participle of regular verbs'; this definition does not give any semantic information but rather informs the user about the grammatical characteristics of the suffix. The suffix *-ization* is an undefined run-on in the entry for the suffix *-ize*, whose meaning is divided into three senses: 'to become, make or make like', 'to speak, think, act, treat, etc. in the way mentioned' and 'to place in'. The suffix *-er* when used to form nouns has four senses in OALD8: 'a person or thing that', 'a person or thing has the thing or quality mentioned', 'a person concerned with' and 'a person belonging to'. Were the user to be somehow familiar with these meanings, s/he might not find it too difficult to infer the meaning of, for example, *chain-smoker* ('a person that smokes cigarettes

continuously'), after having studied the meaning of the verb *chain-smoke*, which is defined as 'to smoke cigarettes continuously, lighting the next one from the one you have just smoked'. It is, however, highly unrealistic to expect the user of a monolingual learner's dictionary to be aware that a certain suffix may also convey a certain meaning. Even if the user did know that, it cannot be expected that s/he would be familiar with the meanings of so many different suffixes, especially if we take into account that many suffixes have different senses. It could perhaps be anticipated that the suffix *-ability* would pose fewer problems, but this may be at least partly because, as far as its form is concerned, it is identical with the noun *ability*; consequently, the user who is unfamiliar with this noun can look it up and learn the meaning. Interestingly, the suffix *-ability* and its variant *-ibility* form an entry on their own, but the user is cross-referred to the entry for the suffix *-able*, which is, together with its variant form *-ible*, defined as 'that can or must be' and 'having the quality of'. At the end of the entry for the suffix *-able/-ible*, two undefined run-ons can be found, i.e., *-ability*, *-ibility* and *-ably*, *-ibly*, which means that the suffix *-ability/-ibility* can actually be found in two different places within one dictionary, but without being explained. To sum up, it would be advisable to provide definitions for such nouns instead of just running them on, because the meaning of the noun in question would be clearer and more easily understood by the majority of users.

As has been mentioned in the Results section, nouns with the status of an undefined run-on can also be appended to the homonymous lemma. The user who looks up, for example, *cerise* would learn that this adjective means 'pinkish-red in colour'. Hopefully, s/he would be able to decode the meaning of the nominal undefined run-on *cerise* as 'a pinkish-red colour'. Similarly, the verb *simper* is defined as 'to smile in a silly and annoying way', which means that users should somehow "guess" that the undefined noun *simper* means 'a silly smile'. This may seem an easy task, but many users would most certainly find it daunting, if not impossible.

It is well-known that learner's dictionaries tend to include precise, complete information about the grammar of the lemma. If nominal run-ons are studied more carefully, we could see that dictionary users can also learn something about the grammatical behaviour of these words. The same policy as for the defined lemmata is also employed for the run-ons. If the grammar code in nouns is absent, it means that the noun is countable (e.g., *signer*, *sight-seer*, etc.). Some nouns are followed by examples of use which sometimes make it clear that the noun is countable (e.g., both examples illustrating the use of the noun *signer*, i.e., *the signers of the petition*; *signers communicating information to deaf people*). Run-ons that are uncountable or always used in the singular are equipped with grammar codes: [U] (e.g., *centralization*, *centre-left*, *sincerity*, *sing*, etc.) or [singular] (e.g., *sizzle*). Again, examples of use, if provided, additionally illustrate the grammatical characteristics. At first sight, the policy of providing grammatical information about undefined run-ons seems not to be consistently pursued. The question can be raised whether the user is supposed to know that

countable nouns are not labelled, and another question that can be asked is how and/or where the user can learn this. In the section entitled *Key to dictionary entries* (pp. ix-xi), no explanation is provided, but the user is referred to pages R11-12 in the back matter, where the following information on different types of nouns can be found:

Countable nouns are the most common type of noun. If they have only one meaning, or if all the meanings are countable, they are just marked *noun*. For nouns that have a number of meanings, some of which are countable, each meaning that is countable is marked [C].

It seems overly optimistic to expect an average dictionary user to read the *Reference Section* in order to be able to learn all the details regarding the (grammatical) labelling of lemmata and run-ons. Would it not be more practical and user-friendly to assign each noun a grammatical code? In this case, the user would not have to study the front or back matter to find information about the inclusion or omission of grammatical information. On the other hand, the grammatical labelling of nouns is very detailed, since a distinction is made between uncountable nouns [U] and singular nouns [singular]. Again, the question can be posed whether average dictionary users will be able to understand and tell the difference between both types of nouns, even if they have read the explanation on page R11:

Uncountable nouns [U]: "An uncountable noun has only one form, not a separate singular and plural. It can be used with or without a determiner." [...] "With nouns such as **furniture**, **information** and **equipment**, as with many other uncountable nouns, you can talk about amounts of the thing or separate parts of the thing by using phrases like **a piece of**, **three items of**, **some bits of**. Nouns like **piece**, **item** and **bit** are called PARTITIVES when used in this way."

Singular nouns [singular]: "Some nouns are always singular and have no plural form. Many nouns like this can be used in only a limited number of ways. For example, some singular nouns must be or are often used with a particular determiner in front of them or with a particular preposition after them. The correct determiner or preposition is shown before the definition."

To sum up the discussion about the grammatical labelling of nouns, it would be advisable to use a suitable code to label each noun, regardless of its status (lemma or run-on) and its grammatical characteristics.

Adverbs, especially those formed by adding *-ly* to the related adjective, are one of the commonest types of run-ons (Atkins and Rundell 2008: 237), which is also in line with the results obtained when studying our database (for a more detailed discussion of adverbial run-ons in British monolingual learner's dictionaries cf. Vrbinc and Vrbinc 2013). Atkins and Rundell (2008: 237) warn that care needs to be taken when some adverbs in *-ly* are used as intensifiers (no such example can be found in our database), or when some adverbs have a double function: manner adverbs (e.g., *reply in characteristically robust style*) and sentence adverbs (e.g., *Characteristically, Helen paid for everyone*). They recom-

mend that in learner's dictionaries, it is best to avoid anything other than simple manner adverbs in the run-on slot, since such adverbs would require two separate senses: one sense for the manner adverb and one for the sentence adverb. As an example of such a treatment that would also be advisable in monolingual learner's dictionaries, we can take the two definitions for *characteristically* from *Collins English Dictionary* online:

1. in accordance with the nature of the person or thing involved, typically ⇒ *Hayes was characteristically modest about his achievement.* ⇒ *He replied in characteristically robust style.* ⇒ *a characteristically polished performance*
2. as was characteristic of him/her/it; predictably ⇒ *Characteristically, Puccini was dissatisfied with the end result.* ⇒ *Characteristically, he worked hard at the assignment.* ⇒ *She knew she was indeed dying; characteristically she did not tell her fellow actors.*

The third most frequent part of speech among the run-ons included in our database is the adjective. The users are doubtlessly faced with a dilemma in decoding the meaning of the run-on adjective because they are supposed to be familiar with the meaning(s) of individual suffixes. The suffix *-less*, for example, could perhaps be considered more transparent as regards its meaning ('without') than suffixes such as *-ic* ('connected with'), *-y* ('full of; having the quality of'), *-ous* ('having the nature or quality of'), etc. It is, however, not to be expected that the average users will first recognize the suffix, then look up the suffix in the dictionary and study its meaning and finally, decode the meaning of the adjective with a particular suffix. For this reason, it would be advisable to define such adjectives rather than just run them on. For example, *slanderous* can be found under the noun *slander*, which is defined as 'a false spoken statement intended to damage the good opinion people have of sb; the legal offence of making this kind of statement', which means that the user is supposed to decipher the meaning of the adjective *slanderous* from the definition for the noun *slander* and the definition for the suffix *-ous*. It would, however, be more user-friendly to define *slanderous* as 'untrue and intended to damage the reputation of the person that it refers to', thus enabling the user to get comprehensive information in one place rather than looking up the noun and the suffix and trying to construe the meaning.

Although included as undefined run-ons, adjectives may also have a grammatical label in order to show users how to use them correctly (cf. also *Reference Section* in the back matter of OALD8, p. R12). For example, the adjective *slighting* has the label [only before noun], which clearly tells users that the adjective is only used attributively.

Verbs are only exceptionally included as run-on entries — only six were found in the stretches studied for the purpose of our research. The question that should be addressed is why these verbs are not awarded full entry status. The answer may be sought in the frequency of these verbs. The compilers of

modern dictionaries, including monolingual learner's dictionaries, claim that dictionaries rely on various corpora. OALD8 is no exception, and on the back cover, we learn that it is "based on the authority of the British National Corpus and the Oxford English Corpus". It is generally known that a well-balanced corpus shows a great variety of contexts, thus providing a good source of definitions for most words. Apart from that, a corpus offers numerous examples of a particular usage and supplies the compilers with information about the frequency of the word in question (cf. also Landau 2001: 296-297). Does this mean that by measuring relative frequency, the corpus told the compilers of OALD8 that the above-mentioned verbs were not worth being treated as lemmata? If we check the frequency of verbs included as run-ons and compare it with the frequency of the nouns to which the verbs are run on, we get the following results:

	BNC (number of hits)		ukWaC (number of hits)	
	noun	verb	noun	verb
catnap	4	5	31	33
caw	3	2	121	22
checkmate	15	13	159	71
sideswipe	8	2	95	26
sightseeing	184	/	4,073	/
sightsee	/	14	/	97
silk screen	0	0	0	0
silk-screen	2	0	53	0

**Table 4:** Number of hits in BNC and ukWaC

As is evident from Table 4, frequency was clearly the criterion, since only in *catnap* is the number of hits for the verb slightly higher than for the noun. In all other cases, the number of hits for the noun exceeds that for the verb. It is, however, true that in *catnap*, the number of hits in both corpora is practically the same for the noun as for the verb, and the same holds true of the number of hits for *caw* and *checkmate* in the BNC. In the ukWaC, the number of hits for the noun and the verb *caw* and *checkmate* is distinctly different. A great difference in the number of hits in both corpora can also be observed in the case of *sideswipe* as well as *sightseeing* vs. *sightsee*. As far as the spelling of *silk-screen* is concerned, it can be seen that both corpora include the hyphenated spelling only, and they both agree that this lexical item is a noun and not a verb. Although frequency seems to play an important role, we should be aware of the fact that a corpus is a source of evidence rather than a source for every single decision made by a lexicographer. A corpus is essential in guiding the compiler to make important decisions, but eventually it is the compiler who needs to take decisions that satisfy the needs of potential dictionary users to the greatest extent possible.

As far as the provision of grammatical information for verbs is concerned, users are supposed to study the *Reference Section* in the back matter of the dic-

tionary (pp. R5-R8). In this section, they learn that the codes [T] and [I] are used to refer to transitive and intransitive verbs, but if a verb is always transitive in all its meanings, it is just marked *verb*. Here, we can draw a parallel with the labelling of nouns, where countable nouns are also not labelled if they are countable in all senses. However, it would be advisable for the sake of consistency to label all the verbs [T] and [I], irrespective of whether they may be transitive in all senses. This does not hold true just of verbs treated as entries but also of verbs included as run-ons. Sometimes other labels are added to draw users' attention to a certain peculiarity, e.g., besides including the grammatical code [intransitive], the verb *sightsee* also includes the label (only used in the progressive tenses). Since transitive verbs can take different types of object, the different patterns or verb frames are shown in bold type, usually just before an example showing that pattern in context, e.g., the verbs *checkmate* and *sideswipe*, which can be used in the pattern ~ **sb/sth**, or the verb *silk-screen* in the pattern ~ **sth**.

As has been established in the Results section, the examples of use are provided for 57 out of 145 undefined run-ons, i.e., 39.3%. Generally speaking, the basic criteria that should be satisfied if an example is to be described as a good example are naturalness and typicality, informativeness and intelligibility, and the right balance between these three criteria is of great importance if the examples of use included in the dictionary are to bring real benefits for the users (Atkins and Rundell 2008: 458-461). We should now examine the run-ons whose use is exemplified to see whether the examples meet these criteria and whether they help the reader either to better understand the meaning of the run-on or to use the run-on correctly.

Doubtlessly, an example should clearly show the context in which a lemma or a particular sense of the lemma typically occurs, which also holds true for examples illustrating the meaning(s) of run-ons. These examples should be chosen with the utmost care, since run-ons lack definitions, and users can only rely on examples if these have been provided. Even words that seem not to pose problems, being regarded as straightforward, may, in fact, not be so straightforward to a foreign learner and dictionary user who may face challenges ranging from problems with grammar to problems with collocations. From this point of view, some examples of use seem not to meet these criteria:

*side-splittingly funny* (used to illustrate the use of the undefined adverb *side-splittingly*);

*a sideways move* (used to illustrate the use of the undefined adjective *sideways*);

*a vow of celibacy* (used to illustrate the use of the undefined noun *celibacy*);

*ceramic tiles* (used to illustrate the use of the undefined adjective *ceramic*).

Because these examples are used in truncated form and lack context, users do not get any information whatsoever about how or in what kind of context to

use the run-ons correctly. Also, the examples that are supposed to illustrate the meaning of agent nouns probably do not make much sense, since they contribute neither to the understanding nor to the correct use of the words in question. For example:

*He's an Olympic silver medallist* (used to illustrate the use of the undefined noun *silver medallist*).

Since users aim to construct English sentences that are syntactically, colligationally and collocationally natural, the illustrative examples should incorporate as many pieces of information as possible. Besides that, it should be clear from the illustrative examples what the word defined really means. This can be illustrated by the following examples taken from our database:

*The game will be broadcast simultaneously on TV and radio* (used to illustrate the use of the undefined adverb *simultaneously*);

*The hotel is centrally located for all major attractions* (used to illustrate the use of the undefined adverb *centrally*).

In some examples, part of the example is printed in bold, which should immediately catch the users' attention. The bolded part of the example indicates a collocation (cf. also p. R13 in the *Reference Section* of OALD8, where it is explained that important collocations are printed in bold type within the examples). For example:

*a **cheerleading** squad/team* (used to illustrate the use of the undefined noun *cheerleading*);

*I can say **in all sincerity** that I knew nothing of these plans* (used to illustrate the use of the undefined noun *sincerity*).

Here, a parallel can be drawn between the examples for *cheerleading* and the truncated examples mentioned above. Although collocations are useful for a foreign learner, it would be more sensible to contextualize them rather than list bare collocations only.

If the compilers of OALD8 believe that the meaning of the example is not obvious, a short explanation follows in brackets. Some examples of this kind can also be found in our database:

*Is the house **centrally heated*** (= does it have central heating) (used to illustrate the use of the undefined adverb *centrally*);

*I could **cheerfully** have killed him when he said that* (= I would have liked to);  
*She **cheerfully** admitted that she had no experience at all* (= she wasn't afraid to do so) (both examples illustrate the use of the undefined adverb *cheerfully*).

As Atkins and Rundell (2008: 460-461) state, the information in the example

should not be in conflict with what the definition says. The adjective *cheerful* is defined as '1. happy, and showing it by the way that you behave 2. giving you a feeling of happiness'. From this point of view, the example *I could cheerfully have killed him when he said that* seems to contradict the definitions, because *cheerfully*, which is supposed to imply something positive, is used with *kill*, which has a negative connotation. Atkins and Rundell (ibid: 461) rightly point out that such an example would be "a perfectly natural thing to say in the right context, but for a learner who has struggled to process the definition, and who believes s/he has grasped the concept, it can only be discouraging to find an example that seems to contradict all this".

As has been mentioned, intelligibility is one of the criteria required to make a good example. This means that, although an example is authentic, natural and typical, it may fail to be informative because the user is unable to understand it. Consequently, such an example is of no use or value to the user (Atkins and Rundell 2008: 461). An example taken from our database can best illustrate this:

*Cheerleading used to be peachy blondes waving pompoms at football games* (used to illustrate the use of the undefined noun *cheerleading*).

The words that many users would probably find incomprehensible are *peachy* (yielding 23 hits in the BNC, but these hits refer mostly to sense 1 in OALD8, i.e., 'like a peach in colour or appearance', and not to sense 2, i.e., '(NAmE, informal) fine; very nice') and *pompom* (yielding only 2 hits in the BNC). Apart from that, this example is also sexist and dependent on outdated cultural information.

Another example that is worth mentioning is the undefined noun *skewbald*, which is exemplified by the sentence *He was riding a skewbald*. It is questionable whether the dictionary user will be able to decode the meaning of this noun by simply reading the definition of the associated adjective *skewbald*, i.e., '(of a horse) with areas on it of white and another colour, usually not black'. It can be claimed with a high degree of certainty that the majority of users will not pay particular attention to the sense indicator provided in brackets, thus failing to understand that the noun *skewbald* actually means 'a horse or pony with patches of white and another colour, but usually not black'.

Another issue that should be addressed is the problem of a polysemous entry at the end of which an undefined run-on can be found whose use is illustrated by means of an example of use or several examples of use. In OALD8, the adjective *central* has five senses: 'most important', 'having power or control over other parts', 'in the centre of an area or object', 'easily reached from many areas', '(of a vowel) produced with the centre of the tongue in a higher position than the front or the back'. The adverb *centrally* is an undefined run-on, whose use is illustrated by three examples, i.e., *The hotel is centrally located for all major attractions*, *a centrally planned economy* and *Is the house centrally heated* (= does it have central heating)? It is evident from these three examples that the

meaning of *centrally* can be surmised in the first two examples if the user is familiar with the meanings of the related adjective but not in the third example, where *centrally* is explained by 'does it have central heating'. In this example, the dictionary user needs to know the meaning of the compound 'central heating' (which is an entry in its own right in OALD8), rather than any of the meanings of the adjective *central*. Since the derived adverb is placed at the end of the entry for the root word, which may have more than one sense, another question should be raised, i.e., to which of the root word's senses does the derived adverb belong? Does it belong to all of them or just to one of them? In some cases, it may be easier to connect the individual examples illustrating the use of run-ons with a particular sense of the lemma at the end of which the run-on is listed. This is mostly the case if a polysemous lemma has fewer senses, and if the undefined run-on contains examples illustrating the use of all senses of the lemma. For instance, let us consider the two examples found under the undefined verb *checkmate*. It can be seen that the example *His king had been checkmated* is linked to sense 1 of the noun *checkmate*, i.e., '(in chess) a position in which one player cannot prevent his or her king (= the most important piece) being captured and therefore loses the game', whereas the example *She hoped the plan would checkmate her opponents* is linked to sense 2 of the noun *checkmate*, i.e., 'a situation in which sb has been completely defeated'.

Quite rightly, the question should be raised how anyone can expect an average dictionary user to know whether the meaning of a particular run-on as suggested in the examples is semantically derived from the first, second, third or other definition of the lemma, or perhaps from all of them. Since OALD8 is intended for learners of English, such treatment of undefined run-ons should be considered undesirable because it is not to be expected that non-native speakers of English would be able to draw any parallels between the meaning(s) of the lemma and the run-on that is derived from it. In all such cases, it would be much better to include the run-on as the lemma and define all its senses separately. If, on the other hand, compilers of a dictionary have to include as many items as possible and if they have to save space at the same time, they necessarily resort to undefined run-on entries. This being the case, the following should be observed: only those words can be included as undefined run-ons whose meanings are truly transparent and easily inferred from the meanings of root words. This means that derived words that have acquired senses not adequately covered by the root words to which they are run on should be included only as main entries (e.g., *centrally heated*). Another solution to the problem is a practice employed by COBUILD7. This dictionary treats undefined run-ons in a more user-friendly way, since they are listed at the end of each individual sense of the lemma, together with at least one example of use. Such a treatment makes it easier for the user to establish a connection between the meaning of the lemma and the run-on.

As regards the inclusion of run-ons, one of the most obvious problems users may face is disruption of correct alphabetical order. For example, *cheerily*

is included under the adjective *cheery*, *silkinness* and *silkiy* under the adjective *silky*, so that alphabetical order is clearly disrupted. Strict alphabetical order is breached by the nesting of derivatives at the end of a certain entry, a practice which may not be transparent to many users (Hartmann 2001: 64). Such treatment may lead to more or less serious difficulties with the look-up process in print dictionaries (the electronic versions do not pose such problems). It should, however, be stressed that the success of any look-up operation may lie in the degree of the user's ability and knowledge about how to retrieve different pieces of information from a dictionary.

The last remark that should be made regarding the material collected for the purpose of our study concerns the undefined noun *sledding*, which is run on to *sled*. *Sled* is labelled with a multiple PoS label, i.e., noun, verb, and has a cross reference to *sledge*. This means that the user cannot possibly decipher the meaning of *sledding* because s/he cannot find any definition for *sled*. In the entries for the noun and the verb *sledge*, *sledding* is not mentioned. The best solution would probably be to include *sledding* as a lemma and provide a definition for it. Another such example is the undefined noun *Catholicism*, with a cross reference to *Roman Catholicism*, which appears at the end of the entry for the noun *Catholic*, where the user is cross-referred to *Roman Catholic*. In cases where a user is cross-referred from one entry to another without being able to find any definition for the particular word, the question can be asked what is actually the point of including such run-ons in a dictionary. It should, however, not be forgotten that OALD8 belongs to the category of dictionaries intended for foreign learners; thus, the inclusion and treatment of various types of information should be carefully considered to make them as user-friendly as possible.

## 5. Conclusion

Theoretically, the meaning of undefined run-ons should be decoded in two ways: with the help of the definitions for the lemma from which the run-on is derived, and with the help of a combination of these definitions and the definitions for the suffix (if the suffix has a semantic meaning). These two methods are problematic in several respects. Firstly, the polysemous lemma and its run-on do not necessarily coincide in all the senses; consequently, the user cannot be supposed to know which senses of the lemma s/he can apply to the run-on. Secondly, the semantic meaning of the suffix is very vague; thus, the definitions are difficult to understand, let alone to apply to the combination base + suffix. Most importantly, we should not forget that OALD8 belongs to the group of learner's dictionaries, which means that the users' native language is not English. This means that the entry should be tailored in a way that is most suitable for the intended users, that is foreign learners. Even if lexicographers provide one or more examples illustrating the use and/or context of the run-on, this does not necessarily help the user to decode the meaning of the run-on.

These problems can be solved in two ways: either by listing the undefined run-ons at the end of each individual sense of the lemma, together with at least one example of use (this method is employed by COBUILD7), or by including and treating polysemous derivatives as entries in their own right and not as run-ons. Finally, it should be clear to the user when and why a certain word is included as a run-on and not as a lemma, and how s/he can establish a relation between the meaning of the lemma and a run-on that is listed at the end of the entry for the lemma.

In print dictionaries, space saving has been a convenient "excuse" for various more or less user-friendly practices for decades — one of these being the inclusion of run-ons, whose main aim was to achieve broader coverage at a low cost in terms of space. This has become rather outdated in the era of modern technology, when even lexicographers often express doubts about the future of print dictionaries (Macmillan, for example, announced in November 2012 that, in future, only online dictionaries will be available), and if print dictionaries are a thing of the past, space-saving methods need no longer be employed. However, it should be emphasized that print dictionaries are still an entirely dominant force in many developing countries and will probably remain so for many years to come.

To conclude, we should bear in mind that dictionary users who look up an unknown or unfamiliar word in a dictionary expect to find the meaning of this word. If the word is included as an undefined run-on, they may still fail to understand it and may consequently be unable to use it. This is also in line with the observations made by Bergenholtz and Gouws (2013: 73, 75), who rightly speak in favour of a presentation and treatment determined by the users' needs.

## Bibliography

### Dictionaries

**Carroll, Katherine (Ed.).** 2012<sup>7</sup>. *Collins COBUILD Advanced Dictionary of English*. Glasgow: HarperCollins. (COBUILD7)

*Collins English Dictionary*. Available at <http://www.collinsdictionary.com/dictionary/english>.

**Turnbull, Joanna (Ed.).** 2010<sup>8</sup>. *Oxford Advanced Learner's Dictionary of Current English*. Oxford: Oxford University Press. (OALD8)

### Other literature

**Atkins, B.T. Sue and Michael Rundell.** 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.

**Béjoint, Henri.** 2010. *The Lexicography of English*. Oxford: Oxford University Press.

**Bergenholtz, Henning and Rufus H. Gouws.** 2013. The Presentation of Word Formation in General Monolingual Dictionaries. *Lexikos* 23: 59-76.

- Bukowska, Agnieszka Anuszk.** 2010. Sampling Techniques in Metalexigraphic Research. Dykstra, Anne and Tanneke Schoonheim (Eds.). 2010. *Proceedings of the XIV EURALEX International Congress*: 1258-1269. Ljouwert: Afûk.
- Hartmann, Reinhard R.K.** 2001. *Teaching and Researching Lexicography*. Harlow: Pearson Education.
- Hartmann, Reinhard R.K. and Gregory James.** 1998. *Dictionary of Lexicography*. London/New York: Routledge.
- <http://www.macmillandictionaryblog.com/bye-print-dictionary>
- Landau, Sidney I.** 2001<sup>2</sup>. *Dictionaries: The Art and Craft of Lexicography*. New York/Cambridge: Cambridge University Press.
- 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. Available at <[https://repozytorium.amu.edu.pl/jspui/bitstream/10593/799/1/Lew\\_space\\_restrictions\\_in\\_paper\\_and\\_electronic\\_dictionaries.pdf](https://repozytorium.amu.edu.pl/jspui/bitstream/10593/799/1/Lew_space_restrictions_in_paper_and_electronic_dictionaries.pdf)>. [Accessed 15 May 2014.]
- Stein, Gabriele.** 2002. *Better Words: Evaluating Efl Dictionaries*. Exeter: University of Exeter Press.
- Svensén, Bo.** 2009. *A Handbook of Lexicography: The Theory and Practice of Dictionary-making*. Cambridge: Cambridge University Press.
- Vrbinc, Alenka and Marjeta Vrbinc.** 2013. The Status of Adverbs in British Monolingual Learner's Dictionaries: Lemmata or Run-ons. *Poznań Studies in Contemporary Linguistics* 49(2): 261-274.

## Corpora

British National Corpus

[https://ca.sketchengine.co.uk/bonito/run.cgi/first\\_form?corpname=preloaded/bnc2;](https://ca.sketchengine.co.uk/bonito/run.cgi/first_form?corpname=preloaded/bnc2;)

ukWaC [https://ca.sketchengine.co.uk/bonito/run.cgi/first\\_form?corpname=preloaded/ukwac3;](https://ca.sketchengine.co.uk/bonito/run.cgi/first_form?corpname=preloaded/ukwac3;)