# Dušan Gabrovšek
University of Ljubljana
Slovenia

# Extending Binary Collocations: (Lexicographical) Implications of Going beyond the Prototypical a – b

## Summary

The paper focuses primarily on the Sinclairian concept of extended units of meaning in general and on extended collocations in particular, investigating their nature and types. Such extended units are extremely varied and diverse; they are regarded as instances of the functioning of the coselection principle. Some extended forms are used far more commonly that the corresponding prototypical (binary) sequences. The final section delves into the ABCs of extended collocations in the context of lexicography, suggesting that dictionaries should make an effort to include a selection of such strings, especially for encoding tasks that are to be shown as examples of use. Most dictionaries incorporate very few such "loose" units, probably because of a powerful tradition to include as examples of use chiefly binary collocations and full sentences.

Key words: extended unit of meaning, extended collocation, pattern, general monolingual dictionary, general bilingual dictionary

# Širitev binarnih kolokacij: (Leksikografske) posledice širjenja prototipičnih dvočlenskih struktur

## Povzetek

Prispevek se osredotoča na razširjene pomenske enote na splošno in zlasti razširjene kolokacije. Ukvarja se z naravo in vrstami tovrstnih enot, ki so izredno raznolike in kompleksne, in sicer v luči t.i. principa besedne zveze Johna Sinclairja. V nekaterih primerih so razširjene enote celó pogostejše kot ustrezne osnovne dvočlenske enote. Zadnje poglavje postavlja osnove za obravnavo razširjenih kolokacij v slovarjih; slovarji bi vsekakor morali vključevati tovrstne enote, predvsem za potrebe uvezovanja, navajali pa naj bi jih dosledno kot primere rabe. Vendar pa večina slovarjev vključuje zelo malo takšnih "ohlapnih" enot; razlog za to je verjetno dolga tradicija, ki med primere rabe v slovarjih uvršča predvsem prototipične (dvočlenske) kolokacije in cele stavke.

Ključne besede: razširjena pomenska enota, razširjena kolokacija, vzorec, splošni enojezični slovar, splošni dvojezični slovar

# Extending Binary Collocations: (Lexicographical) Implications of Going beyond the Prototypical a – b

## 1. The Core Units: Collocation and Extended Collocation

Since about the mid-to-late 1980s, collocations have been a hot topic in English phraseology, in both research and language teaching. Commonly defined as frequent, semantically transparent, recurrent, psychologically salient, typically binary sequences such as *fatal accident, scheduled flight, readily available, to somebody's advantage*, and *to make a statement*, collocations display a variety of combinability restrictions that are often language-specific and as such largely unpredictable in cross-linguistic terms, particularly in encoding-oriented (L1→L2) language tasks. Moreover, a number of collocations allow what Sinclair (1991, 111) labeled "**internal lexical variation**" (Partington 1998, 26), as for example *in some cases/instances*, being indicative of synonymous collocational variation,[1] and also observed e.g. in *to give/make/deliver a speech, at top/full speed, a highly/largely speculative theory, absolute/complete/dead/deadly/total/utter silence* etc. Another feature of a number of even highly restricted binary collocations such as *vested interest(s)* is that they can have words inserted into them to produce **extended syntagms**, or more specifically **extended collocations**, e.g. *vested financial interest* or *vested political interests* (Partington 1998, 26). Considered from a wider phraseological angle that goes beyond collocation alone, this process illustrates the sometimes multistage "buildup-type" creation of a great variety of "extended" sequences that are, for the most part, also (most) commonly used in their basic, prototypical (=mostly binary) forms.[2] This means that the extension process can be observed not only in collocations, where extended collocations are thus duly observed, but also (though less commonly) in **extended idioms** and in **extended compounds**; however, the latter two are indicative of a (somewhat) different process of extension. Overall, this is an important phenomenon in phraseology that may not (yet) have been given its due; if anything, it has important implications for dictionary making, both monolingual and bilingual. Some of the major bilingual lexicographical implications are to be also addressed in the final sections of this paper, chiefly in the restricted framework of *extended collocability*.

## 2. Background: Extended Units of Meaning

After years of dedicated corpus-based research, John Sinclair (1996 [and 1998]; quoted in Stubbs 2002, 63) insightfully suggested that meaning, rather than being primarily a single-word-based feature (the traditional way of looking at it), is to be sought basically in what he dubbed **extended units of meaning**,[3] the underlying idea being that units of meaning are "largely phrasal." Since

---

[1]   I have argued that in an advanced learners' dictionary the logically unassailable principle of "the more the better" need not apply as a matter of course when it comes to synonymous collocations listed in it (Gabrovšek 2011).

[2]   An exception can be found e.g. in *to take a turn*, a binary collocation that is uncommon on its own, being as it is chiefly used in extended sequences such as *to take a sudden/dramatic/new/strange turn* and *to take a turn for the better/worse*.

[3]   The **extended unit of meaning**, first suggested in the 1990s, stands in sharp contrast to most earlier views of the nature of the unit of meaning, covering a complex inventory of units and unit-like sequences whose formation can be viewed as a "buildup" process, as in (two-"level") the prepositional phrase *on hold* and *to put something on hold* ('to delay'); or (three- or four-"level") the noun *eye*, the collocation *naked eye*, the colligational string *to/with the naked eye*, and the extended unit *visible to/with the naked eye* that can itself be expanded to the four-"level" *barely visible to/with the naked eye*. An example with a weaker real-language evidence of the "mid-level" position can be found in *an eye – a blind eye – to turn a blind eye* [to something]. The basic distinction is often one between a collocation or compound and its pattern-type expansion, e.g. *common ground* vs. *to be* [quantifier] *common ground + between*, or *an easy victim – something makes someone an easy victim (for somebody else)*, which contrasts with the fixed but not pattern-type *fall victim (to something)*. A four-"level" example with additional grammatical requirements can be

these units are complex, they have been defined in Sinclairian terms with reference to four lines of work:

a) identification of **collocational profile** (the item and typical collocations it forms)
b) studying **colligational patterns** (the item and its recurrent grammatical environment)
c) establishing **semantic preference** (i.e. identification of a common and well-defined semantic field)
d) possible existence of **semantic prosodies** (these exist in certain collocations only, being in essence semantic "coloring" of the sense of a collocational component).

Stubbs (2004, 122) notes that a – d are increasingly abstract relations, adding that we can also specify another three relevant properties:

e) **strength of attraction** between node and collocates
f) **position of node and collocate**, variable or fixed (as in *spick and span*, but not *\*span and spick*)
g) **distribution**, wide occurrence in general English or in broad varieties (e.g. journalism), or restricted to specialized text-types (e.g. recipes: *finely chopped*, or weather forecasts: *warm front*).

However that may be, one should perhaps treat separately composite "frame"-like combinations, sometimes dubbed **patterns**, i.e. strings with "real words" and diverse "slots". Thus e.g. at the basic (prototypical), i.e. binary, level, while the adjective *devoid* combines (to form a grammatical collocation) with the following preposition *of*, it is – following a buildup process – usually used in a wider pattern formalized in the following way:

> [inanimate noun] + *to be devoid of* + [abstract noun denoting a good quality such as *warmth, compassion, humor, feeling*].

> Likewise, the noun *matter* is typically used, in one of its senses, in the pattern *to be + (just/only) a matter + of + noun* or + *of + -ing*-clause.

Also, the combination *to speak of*, in addition to what it does in syntactic terms, can be used idiomatically with a preceding negative to mean 'very little of something or a very small thing' (Mayor 2009, 1688), as in *At this stage, the young bird doesn't have any wings to speak of.*

Why, one might wonder, do we need extended units of meaning, or unit-like sequences? Why not use, for the most part, single-word items used merely in accordance with grammar rules? The reason is likely to be the **complexity** of what we wish to talk or write about, probably coupled with language economy, resulting in different degrees of not only syntactic complexity but also **phraseological complexity**. This complexity appears to be at its strongest with (fairly) frequent, contextually significant, salient and/or topical issues characterizing our everyday communication.

# 3. Extended Units of Meaning and the Coselection Principle

Extended units of meaning represent one "direction" of the functioning of the **coselection principle**, aka **phraseological tendency** (Sinclair 1991, 110ff., and Sinclair 1996), which operates in a great

observed in the noun *table*, the collocation *conference table*, the prepositional phrase *at the conference table*, and the extended unit *to sit at the conference table*, where *sit* is often in the progressive form and preceded by a noun/pronoun subject. The collocational "buildup" can likewise be observed e.g. in *career – a promising career – a promising career derailed (by alcohol abuse)*, or in *problem – a cocaine problem – a cocaine problem plagued baseball (in the 1990s) – a rampant cocaine problem plagued baseball (in the 1990s)*.

variety of ways; it is at work **whenever the choice of one word affects the choice of others in its vicinity**. We only need to consider, by way of exemplification, any out of a number of conventionalized, "frozen" phrases such as the laudatory SI *le tako naprej!* and its English equivalent, *keep up the good work!* As Lord (1994, 78) puts it, a word is able, to a considerable extent, to "predict" its environment, owing to a strong cohesive tendency between words. Not <u>all</u> words, witness e.g. function words such as *for* or delexical(ized) verbs such as *to take*.

The coselection principle sometimes shows (subtle) semantic changes in superficially similar binary and extended sequences that may be of relevance in foreign language teaching; thus many English teachers are well aware of the possibility of their students' misinterpreting or confusing the meaning of *to be sure of (doing) something* ('to be certain to get something or that something will happen') vs. *to be sure to do something* ('remember to do something'), or failing to grasp the role of the article and thus confuse or regard as synonymous *to take the place (of)* and *to take place*. Also, many learners are blissfully unaware of the fact that *at the beginning* is usually followed by a prepositional phrase starting with *of*, whereas *in the beginning* is usually used on its own. On the other hand, phraseology may be responsible for semantic distinctions that do not apply to the otherwise synonymous single-word items making them up; thus e.g. the prepositions *below* and *under* are more or less synonymous as single-word items, whereas *below the belt* is an idiomatic expression meaning 'unfair, cruel' (a comment may *hit below the belt*), whereas *under your belt* means '[having something] useful or important' (an employee with several years' experience *under his belt*).[4]

Even some of the more recent grammars (sic) have demonstrated an increased awareness of the many-sided nature, and pervasiveness, of the functioning of the coselection principle, witness e.g. the substantial treatment, in the corpus-based *Longman Grammar of Spoken and Written English* (Biber et al. 1999, 990-1024), of **lexical bundles**, defined as "sequences of word forms that commonly go together in natural discourse" (ibid., 990), i.e. common and structurally often not complete recurrent expressions (such as *going to be a, do you want me to, it should be noted that*) that show a statistical tendency to co-occur.[5] Such sequences could well be regarded as a type of "**extended collocations**" (ibid., 989). However, in this paper, an *extended collocation* (alternatively referred to as **composite collocation**) is a structurally <u>complete</u> sequence in which the prototypical, i.e. binary, collocation – usually regarded as the basic string – has been augmented by at least one lexical item (this is a quite common process) that is largely (con)text-bound; while keeping (either most or some of) its collocational character, each extended sequence is typically (but not unavoidably!)[6] less frequent and phraseologically "looser" than the corresponding basic form, as for example in

> *an embarrassing decision* → *an acutely embarrassing decision,*
> *to give a sigh* → *to give a grim sigh*
> *to make allegations* → *to make serious/false allegations,*
> *to provide an analysis* → *to provide an in-depth analysis,*
> *to receive coverage* → *to receive wide coverage* → *to receive wide(spread) media coverage*

---

[4] Semantic distinction can be occasionally observed even in denotatively synonymous compounds (themselves not very frequent to begin with), witness e.g. *a senior moment*, 'a time when you cannot remember something, because you are getting older – used humorously' and its synonym *brainfart*, which is stylistically marked, viz. informal (Mayor 2009, 1585, 190).

[5] Following the publication of the *Longman Grammar*, the *lexical bundle* was subjected to a typological investigation (e.g. Biber et al. 2003, Granger and Paquot 2008). In Biber and Barbieri (2007, 264), *lexical bundles* are defined as "multiword sequences that occur most commonly in a given register", "the most frequently recurring sequences of words (e.g. *I don't know if*)" that are "usually not structurally complete and not idiomatic in meaning, but they serve important discourse functions in both spoken and written texts".

[6] Thus it is difficult to maintain e.g. that *to take responsibility* just has to be commoner than *to take full responsibility*.

*an allergic reaction → to have/suffer an allergic reaction → to have/suffer a severe allergic reaction*[7]
*to pass a bill → to pass a landmark bill → to pass a landmark health-care bill → to pass a landmark health-care reform bill*
*to score a goal → to score an away goal,*
*to take a break → to take a short/lunch/career break,*
*to take a turn → to take a wrong turn.*[8]

In consequence, most of such extended formations can logically be broken down into two or more collocations making them up (e.g. *to provide an analysis + an in-depth analysis* and *to receive coverage + wide coverage + media coverage* in two of the examples just cited). Let us merely note that aside from "standard" extended collocations, there are in existence also "**patterned**" **extended collocations** that consist of "real" words and "slots" (e.g. *common knowledge → it + to be + common knowledge + that*-clause; *naked eye → to be + [barely] + visible + to/with + the naked eye*).

Incidentally, the extension of a multiword collocation-type string can also be found in another, idiom-type phraseological mode, as "a common type of idiom variation", "making idioms more specific by the addition of words that link them to the context," producing e.g. *toe the education authority line* from the more general *toe the line* (Bernardini 2004, 18).

# 4. Extended Units of Meaning at Large

First of all, we need to zero in on the broad-based **extended units of meaning** at large, such as Sinclair's own oft-cited *(barely) visible to/with the naked eye* referred to above. How exactly does one go about determining them, say in the case of [somebody] *never appeared truly in danger of coming out on the wrong end against* [an opponent], found online (Yahoo) on 26 May 2009 in a daily Roland Garros tennis report? There is more – for instance, combinations such as *think nothing of* (+*ing*-verb), with its contextual, or pattern-type, colligational requirement.

To start with the essentials, the basic-level, or prototypical, i.e. binary, collocations, which are quite common, are those which are extended by another collocating item, as found e.g. in

*to make a discovery → to make a significant discovery,*
*to take a turn → to take a subtle turn,*
*a pay cut → a steep pay cut,*
*a twist of fate → a strange twist of fate.*

There are also cases in which sequences verge on idioms, or are at the very least more idiomatic, such as *to take a turn → to take a subtle turn*.

Such extended formations can take not only additional collocational (=lexical) elements but further "logical" colligational (=grammatical) additions, as in *to keep a watchful eye <u>on</u>* [somebody/ something], to *keep a sharp lookout <u>for</u>* [something/somebody], or *to exert an undue / a strong/*

---

[7] One might argue here that the basic collocation is *to have/suffer a reaction* rather than *an allergic reaction*; for such issues, corpora should be consulted to determine frequency figures.

[8] Extended collocability has been interpreted in different ways, e.g. as **collocational cascades**, or "sequences of interlocking items," where collocational patterns extend from a base to a collocator and on again to another base, creating "chains of shared collocates" (Gledhill 2000, 212). Moreover, one can look at the "buildup" process as one starting with a single-word item, say *work* (n.) → *to do work → to do heavy work → to do heavy outdoor work*. Dictionaries typically prefer the binary combination; thus in the *Longman* (Mayor 2009, 817), from which this example was taken (*heavy* adj, sense 3), only *heavy* and *work* are printed in bold type.

*powerful influence <u>on/over</u>* [something/somebody]. Again, such extended units can be **non-collocational**; they can be, though less frequently, either idioms or compounds. But are we to argue that their length aside, extended collocations are fundamentally different from (the decidedly fewer) **extended idioms** including idiomatic phrasal verbs, and **extended compounds**, such as

the idiom *to lend [somebody] an ear → to lend [somebody] a sympathetic ear*,
the idiom *to pack a punch → to pack a hard/hefty/strong punch,*
the idiom (in the form of a phrasal verb) *to pick up → to pick up an accent*,
the compound *the all-clear → to give the all-clear*,
the AmE compound *rain check → to take a rain check → to take a rain check on something,*
the compound *cloud nine → to be on cloud nine,*
the compound *senior moment → to have a senior moment,*
the compound *self-inflicted → a self-inflicted wound → a self-inflicted gunshot wound,*
the compound *self-image → a positive/negative/good/poor self-image,*
the compound *pickup → pickup truck*, and
the compound *self-fulfilling → self-fulfilling prophecy*?

Do extended forms preserve their prototypical, basic-form features? Indeed, quite a few of them merely add (slight) emphasis, so that they are largely synonymous with their prototypical counterparts, as in *to keep an eye out for → to keep a sharp eye out for*. Come to think of it, there are in existence also quite a few **(near-)synonymous extended collocations** (but far fewer synonymous extended compounds and extended idioms), e.g. *to be in need → to be in urgent/desperate need*, or in *to bear a resemblance → to bear a close/great/marked/remarkable/striking/strong/ an uncanny/eerie resemblance to somebody/something* (McIntosh 2009, 693).[9] So why make a fuss over the issue? Not really – matters are not nearly as simple and straightforward.

First of all, interestingly, some of such extended units – whether collocational or non-collocational (=idiom- or compound-type, provided they are considered to be comparable) – appear to exist, for the most part at any rate, chiefly or even solely in the extended form, as in the case of the idiom-type string *to give somebody/something a clean bill of health*, or the "mixed," idiom-cum-collocation-type *to keep a straight face*, where the verb *to keep* virtually always collocates with the entire nominal collocation rather than with the headnoun *face* alone. Next, *to take a twist* is characteristically used in an extended form such as *to take another twist* or *to take a new/cruel/unexpected/strange etc twist*, itself often further expanded into e.g. *to take a deadly new twist*. Also, *long-haul* is chiefly used only attributively, being as it is normally followed by *flight/route/destination* and only a few other similar nouns. Such examples may suggest that collocations can easily "become" larger while still keeping (some of) their phraseological, i.e. collocational, character.

Similarly, also some idioms are commonly used in "logical" colligational extensions, such as *to cut corners <u>on</u> something*. Finally, in certain extended items the extension is not restricted to one item since it offers synonymous alternatives, and in some others the basic (=unextended) phraseological form hardly enjoys phraseological salience, as it were, as in the compound *taskmaster* that is mainly used in *to be a hard/stern/tough taskmaster*, 'to force people to work very hard' (Mayor 2009, 1804).[10]

---

9    Note that the more selective *Longman Collocations Dictionary and Thesaurus* (Mayor 2013, 1039) only offers, for this particular sense, *a close/strong/striking/remarkable resemblance* and *an uncanny resemblance*; still, the idea of (near-)synonymous collocations remains unchallenged. Another collocational competitor, the *Macmillan* (Rundell 2010, 690), shows the same thing in not one but, interestingly, two sections to capture visually the "intensity" distinction in the *resemblance* entry, as follows: *close/distinct/ great/marked/strong resemblance* and *remarkable/startling/striking/uncanny resemblance*.

10   Overall, as the above examples suggest, the complexity of the issue of phraseological extension goes a long way toward indicating

# 5. The Concept of Wider Patterning

Extended units of meaning are clearly related to the – considerably broader – concept of **wider patterning** (Francis 1993; Hunston and Francis 2000, 1-3); in this respect, it has been observed that a word often comes with its "**attendant phraseology**"; for instance, the noun *matter* frequently occurs with '*a ___ of* -ing,' this phraseology here being "the grammar pattern belonging to the word *matter*". The interaction between particular lexical items and grammatical patterns they form a part of is of the utmost importance, going as it does beyond the limits of particular collocations. In this approach, a **pattern** is "a phraseology frequently associated with (a sense of) a word, especially in terms of the prepositions, groups, and clauses that follow the word", or (ibid., 35) "all the words and structures which are regularly associated with the word and which contribute to its meaning." When used in this sense, *pattern* signifies a lot more than frequent co-occurrence.[11] A pattern can be identified if a combination of words occurs relatively frequently, if it is dependent on a particular word choice, and if there is a clear meaning associated with it.

# 6. Extended Collocations and Kinds of Extension

While recognizing the complexities involved in phraseological, or syntagmatic, extensions at large, this paper focuses on extended collocations alone. What must be pointed out at the outset of this section is that extended collocations can be quite predictable in pragmatic terms and thus less collocationally salient, i.e. less phraseological, in terms of the item used for extension. Thus for instance in *to foster creativity* and *to foster a child's creativity* the added noun is far from being phraseologically special, that is, hardly significant in collocational terms.

An important status-related issue is likely to arise when one starts considering the **syntagmatic status** of extended collocations. Is such a string a phraseological unit to begin with? Is it on a par, as it were, with the "basic" collocation, or is it merely an expanded – significant because recurrent – typical co-text of the base, to be regarded perhaps merely as a typical example of use, devoid of any additional specific phraseological features? The issue is usually examined either in a monolingual context, where frequency of co-occurrence typically prevails, or in a bilingual framework, where what really matters is rather overall idiomaticity in the broader sense (i.e. when used to refer to restricted nativelike textual selection and restricted nativelike textual sequencing) as well as the twin contrastive criteria, viz. unpredictability (semantic) and non-congruence (structural) as observed in cross-linguistic phraseology and translation, especially in encoding tasks (L1→L2).

The extended collocation may take a variety of forms aside from the prototypical one, viz. that of a binary lexical collocation adding another lexical element, sometimes with synonymous alternatives, such as

> *war crime* → *war crime suspects,*
> *to give a cry* → *to give a sharp cry,*
> *to make progress* → *to make good progress*, or
> *to take a breath* → *to take a deep/long/big breath.*

---

the diverse ways in which the co-selection principle operates in the billions of acts of communication carried out in everyday language use.

[11] Hunston (2004, 112) argues that grammar information in a learners' dictionary be given in the form of patterns, because they capture what a learner needs to know about a particular word. This is consistent with a *lexically-driven* concept of language.

Another common type of extension involves a colligational constituent and can well start with a single-word item; it can be illustrated e.g. by

*rampage* → *on the rampage* → *to go on the rampage* (where the noun is virtually always to be found only with *on the*, with *go* being very frequent but not indispensable), or
*pressure* → *under pressure* → *to be/come under pressure*.

**Extended forms** are often more complex and/or varied, consisting of e.g. either a phrasal verb + a single-word item or a compound + a single-word item, or some conceivable variation of these, as in

*foul play* → *to suspect foul play*, and *foul play* → *to rule out foul play*,
*lip service* → *to pay lip service to somebody/something*,
*hands down* → *to win (something) hands down*, and *hands down* → *to beat somebody hands down*, or
*second nature* → *to be/become second nature to somebody*.

Intriguing cases include phraseologically different but still basically synonymous sequences such as *an educated guess* (compound) and *an informed guess* (collocation), both of which are commonly extended by the verb *to make* (resulting in *to make an educated/informed guess*), or *to give a performance* that can be extended, in the same sense, by either *terrific* or *stellar* (resulting in *to give a terrific/stellar performance*). Furthermore, as has already been mentioned, an extended form can incorporate not only collocational but also colligational elements, as in *to make a discovery* → *to make a significant discovery* → *to make a significant discovery about something*. Likewise, the extension may transform a lexical collocation into a lexical-grammatical one, or indeed the other way around, another possibility being for a grammatical collocation to be expanded into a lexical-grammatical one, as in

*make an accusation* → *make an accusation against* (someone),
*make an attempt* → *make an attempt to do something / at something / at doing something*,
*give one's approval* → *give one's approval to something*,
*in progress* → *work in progress*,
*be in need of something* → *be in urgent/desperate need of something*.

Clearly, then, an **extended lexicogrammatical unit** may be formed in a variety of ways, some simple and others more complex. For example, it can comprise a compound augmented, in two or more "buildup" stages, either by a single-word item such as *rock bottom* → *to hit/reach rock bottom*, or by a phrasal verb, as in *the nitty-gritty* → *to get down to the nitty-gritty*.[12]

Moreover, the "buildup" process may not only vary in complexity but can, and not infrequently does, consist of more than two stages (cf. footnote 3). Thus in discussing *collocation*, Pearce (2007, 37), for one, points out that a frequent collocate of the noun *gamut* is *whole*, and that the two are typically used in "extended phrases" like *to run the whole gamut*. Likewise, *a look* forms the collocation *to take a look* that can be itself expanded into the three-item *to take a fresh/quick look* and even into the emphatic four-item *to take a long hard look*; moreover, there is often a colligational element (*at somebody/something*) following the string. Also, *faith* forms a frequent collocation, or

---

12   Note that idiomatic-collocational expansions can be phraseologically different from the basic binary item. Thus e.g. the collocation *to take a view* can be seemingly expanded into the idiom-like string *to take the long view*; however, in this case the regularity of the extension may be called into question because of the fixed *the* used idiomatically only in the longer structure.

compound, *good faith* that itself forms part of several frequent and progressively larger – and more complex – combinations: *in good faith, to act in good faith*, and *something is declared as a sign/gesture/ show of good faith*. The extended unit can be a combination of an idiom expanded by a literal single-word element, as in *to mutter* used in front of the idiom *under one's breath* (*to mutter under one's breath*). However that may be, instances of the collocational **"buildup" process** going beyond the basic a - b are not too difficult to find, e.g.

> *problem → the drug problem → to tackle the drug problem,*
> *mind → an open mind → to keep an open mind,*
> *a visit → to make a visit → to make the first visit → to make the first official visit (to).*

Note that combinations that involve phrasal verbs and compounds are not usually counted among extended collocations because most phrasal verbs and compounds are themselves not (regarded as) collocations. Compounds in particular are not commonly regarded as being complex items to begin with.

A caveat: not every (seemingly) expanded string is automatically an extended unit of meaning; those that must be ruled out on logical grounds are either purely grammar-based and -generated, or idiomatic in the sense of being semantically opaque, thus being different from their constituents, as in *to bring up* ('mention,' 'look after,' 'charge with a crime,' 'make something appear on a computer screen') vs. *to bring somebody up short* ('surprise someone and make them stop doing something').

And another caveat: one can sometimes encounter a "pattern-type" extension of a phraseological item, a phenomenon that usually does not get recorded in dictionaries: Thus the complex item *in the way of* or its synonym *by way of*, used to specify the kind of thing one is talking about, is characteristically preceded by a "slot" calling for a restrictive or negative element, as in *… a country without much in the way of natural resources* or *Meetings held today produced little in the way of an agreement* (both taken from the last substantial revision of the *Collins COBUILD* [Sinclair 2001, 1766]). What such "extensions" produce is not to be regarded as extended units, because what is added is not an item as such but rather a **slot-like place** to be filled by *any* out of a number of items fulfilling a general criterion (in this particular case they have to belong to the category of "restrictive or negative elements").

Again, the variety involved in the creation of such composite combinations is great; for example, they can consist of a phrasal verb followed by another single-word lexical item in the singular or the plural, as in *to draw up a list* and *run out of money/ideas*, or of a link verb + phrase, as in *to be up for grabs*. Perhaps the most obvious and "logical" type of the extended collocation is the prototypical, binary collocation extended by the addition of another "pragmatically relevant" lexical element. Such sequences can be made up of a verb + noun with an additional premodifying adjective or attributive noun, as in *to take a turn → to take a wrong turn,* or *to make a visit → to make a surprise visit*.[13] Another commonly encountered possibility comprising a single-word item plus a compound is, one, a nominal compound associated collocationally either with a verb, such as *to do a head count*, or with another noun, the compound being used as an attributive noun, as in *a landmark decision/case* (two lexical elements), or in *to give a keynote speech* (three lexical elements),[14] and two, a compound that exhibits "two-layer" or "two-stage" collocational or colligational links,

---

[13] Note that unlike many of the basic forms, some of the extended collocations may be unexpectedly challenging in the decoding process – and, in quite a few cases, even more so in encoding, thus typically representing encoding problems that can well be nothing short of being formidable.

[14] There is, of course, more variation to be found in premodification: In *to give an after-dinner speech*, the premodifying element is not really a compound but still a complex premodifying element.

as in *fast lane* that is frequently used in the prepositional phrase *in the fast lane* (layer 1) that is itself usually found in *life in the fast lane* (layer 2). Yet another structural possibility is the combination of an adjective or premodifying noun + compound, as in *emotional fallout* or *traffic gridlock*. But then again such combinations cannot be regarded as extended collocations, given that one can hardly consider a compound to be itself a composite unit forming either the node (=base) or the collocator of the extended unit – indeed, the reason why most English dictionaries list compounds as main entries is precisely because their lexical status is on a par with single-word units.

Could one maintain that overall, the resulting extended sequences are phraseologically "looser" and less internally cohesive than their corresponding "tighter" base-level binary units, in a way that might affect their overall lexical status? To a point where their phraseological status might be called into question? Not really. True, one needs to account also for extended collocations that are more syntactic, pattern-like in that they exhibit one or more open "slots" in a "frame," as in *it + to be + common knowledge + that*-clause. Some other combinations may be perceived as being simply longer and thus necessarily more complex than the corresponding "base-level" formations, with syntagmatic preferences being, for the most part, very much in evidence, as in *a long drawn-out war*. A more complex example is provided e.g. by Danielsson (2007, 17-18): The noun *scruff*, 'the nape of the neck,' is not only chiefly used in English in the set phrase *by the scruff of the neck* but also the verbs on the left-hand side of the phrase tend to belong to a group denoting the action of grasping something: *take, grab, drag (in),* or even *pick up*. Moreover, the objects associated with *take by the scruff of the neck* are most often *a game, a match, an opportunity*. Needless to say, such usages are only likely to become obvious when the researcher views the **repetitive patterns** of language data through a concordance program. But what invariably makes them significant as multiword units, or at least unit-like sequences, is in essence their recurrence coupled with the high level of phraseological salience dubbed phraseological experientiality (a concept I suggested and tried to elaborate recently at an EUROPHRAS conference (Gabrovšek 2012)).

It seems only logical that in addition to being investigated within general language, extended collocations and similar units may also be a feature of **specialized discourse** while also showing – keeping – both collocational and colligational links. In many cases, they turn out to be contrastively significant, witness e.g. (from the field of law) the collocational *to plead guilty* and its colligational-cum-collocational extension *to plead guilty to a charge* (SI *priznati krivdo za obtožbo*). Many a non-expert Slovene translator would typically encode the SI structure as *\*to admit guilt for a charge / an accusation*.

In broader terms, too, it has already been implied in this paper that phraseological extension can be seen as a pervasive phenomenon, to be considered in line with Hoey's (2005, 8) suggestion that not only single words, which are acquired through encounters with them in speech and writing, in which process they become **cumulatively loaded** with the contexts and co-texts in which they are encountered, but also word sequences built out of them become **loaded** with the contexts and co-texts in which they occur.[15] Hoey himself goes on to provide the following example to illustrate the process: *winter* collocates with *in*, producing the phrase *in winter*. But this phrase has its own collocations, which are separate from those of its components, so that *in winter* collocates with a number of forms of *to be* (*is, was, are,* etc.), something that neither *in* nor *winter* can apparently do. Similarly, the noun *word* collocates with *say*; the combination *say a word* in turn collocates with *against*, and *say a word against* collocates with *won't*. In this way, a variety of lexical items and bundles are created (ibid., pp. 10-11). But how much of this is to be captured in a (bilingual)

---

[15] This property of sequences has been labeled **nesting**, "where the product of a priming becomes itself primed in ways that do not apply to the individual words making up the combination" (Hoey 2005, 8).

dictionary, and how should one go about assessing the (diminishing) phraseological relevance of such formations? If anything, this kind of analysis does seem to be far more relevant to encoding needs than to decoding needs, and thus to be far more significant in encoding-oriented dictionaries.

# 7. Extended Collocations in the Dictionary: Monolingual and Bilingual

This section concentrates on the ABCs of the treatment of extended collocations in the dictionary, clearly a significant "applied" aspect of the topic. If collocations are largely and uncontroversially an encoding issue, and granted that today's leading British-made advanced monolingual learners' dictionaries of English (notably Mayor 2009 / Fox and Combley 2014, McIntosh 2013, Rundell 2007, and Turnbull and Lea 2010, or their online versions) usually assign collocations and most of the extended collocations that they include the status of **(highlighted) (parts of) illustrative examples of use**, we must conclude at this point that the key question is not necessarily whether this is the best policy in general, but rather whether it is one to be adopted and consistently applied in particular in bilingual dictionaries too. However, one must hasten to add that overall, extended collocations have been given short shrift by most general-purpose dictionaries (including learners' dictionaries of English), the main reason in all probability being that following a long and powerful tradition, English dictionaries strongly favor the inclusion especially of binary-phrase-length and sentence-length examples. Bilingual dictionaries in particular prefer simple (binary) collocations to extended ones, even though not infrequently, this means distorting, in a sense, or perhaps better reducing, the language, or more precisely part of the way we communicate in English, to the basic level. For instance, if, in English, one commonly encounters extended collocational strings such as *to win thunderous applause* (SI *požeti buren aplavz*) and *to take a comfortable lead* (SI *povesti z veliko razliko* or *preiti v prepričljivo vodstvo*), why break such real-language sequences into their binary components as a matter of course?

What at the very outset is a key criterion-related interlingual desideratum is that in a bilingual dictionary an extended collocation should be included – always as (part of) a translated example of use – in the first place if it offers, in translational terms, something more than the binary collocation does. This, to be sure, is a contrastive, or cross-linguistic, consideration. Additionally (or perhaps even initially), there is likely to be also the key monolingual L2 consideration, most of the time at any rate: Ideally, at the same time, extended collocations will be included if they also serve to illustrate idiomaticity of the L2 (English) – idiomaticity here referring broadly (and, unfortunately, rather vaguely) to a sequence exhibiting restricted native-like textual selection and sequencing. What is almost bound to be problematic in devising and implementing a consistent policy of entry selection (inclusion/exclusion) has to do with the overall length of extended collocations to be included in the dictionary. One might well decide to draw, as a matter of principle, on the logical fact that the longer the extended collocation, the less "tightly" phraseological, and indeed the less frequent – hence the more purely contextual – is it likely to be.[16] This means, at least by implication, that the longer the extension, the less likely it is to be a fitting extended collocation to be considered for inclusion in the bilingual dictionary.

[16]    Small wonder, then, that a Google search carried out on 1 April 2012 shows almost 50 million hits for the collocation *the naked eye*, some 10 million hits for *visible to the naked eye*, and just over 2 million for *barely visible to the naked eye*. The figures are comparable to those obtained from a standard corpus, the COCA (http://corpus.byu.edu/coca/), where there are 581 hits for *the naked eye*, 121 for *visible to the naked eye*, and a mere 7 for *barely visible to the naked eye*. However, even the last figure is not really insignificant, if only because there are no hits at all for either ˀ*hardly visible to the naked eye* or *scarcely visible to the naked eye*.

Such being the case, the question must be addressed as to whether one can really frame a sound and consistent dictionary policy (based on both L2 idiomaticity (partly) in terms of the length of the extended collocation and the ensuing L1/L2 translational relationship) regarding the selection and treatment of extended collocations in the general bilingual dictionary. Next, there is the related issue of what inclusion/exclusion criterion or criteria one should apply in the case of the existence of synonymous extended collocations. More significantly, the question of whether or not to draw a principled distinction between "standard" and "pattern-type" extended collocations is to be considered too; it does appear that, overall, extended grammatical collocations with an "added" non-prepositional slot-like element are a lot less phraseological and as such less relevant as a kind of multiword units, witness e.g. the binary lexical collocation *to feel the need* as extended colligationally into the pattern-like sequence of *to feel the need* + *to*-infinitive/infinitival clause. In an advanced learners' English monolingual dictionary, such strings could well be included, albeit selectively, but not (solely) in the form of formalized statements, which is at best artificial to most dictionary users. Since they illustrate – however selectively – typical actual use, they are to be shown primarily as real-life-type language segments, that is, as frequent actual realizations of the pattern-like elements. If they are to be shown in metalinguistic terms as well, the recommended treatment might well be one with concise metalinguistic statements immediately followed by specific "real-life" collocational or colligational extensions. But can – should – this be done in a bilingual encoding-oriented dictionary? Hardly. A judicious selection of translated examples of use appears to be the answer.

# 8. Conclusions

The basic recommendations, then, are the following:

1) A selection of extended collocations should definitely be included (always in the form of (highlighted) (parts of) examples of use) chiefly – but not exclusively – in the general encoding-oriented bilingual dictionary. They can be incorporated either each on its own (extended-phrase-length exemplification) or as (highlighted) parts of sentence-length examples of use. Note that for encoding purposes, one invariably starts with the L1, where the string provided may be distinctly different from its rendering in the L2, and indeed not even be an extended collocation to begin with – or indeed the other way around, witness e.g. SI *popolnoma ozdraveti* EN *to make a full recovery* or SI *pošten/dober spanec*  EN *a good night's sleep*.

2) The selection of extended collocations to be included is to be based on contrastive considerations (translational relevance with respect to the rendering of the basic binary collocation vs. the rendering of the extended one) and, if possible, on the broad-based idiomaticity of the L2 strings being shown as translation equivalents. Frequency[17] is clearly also a factor in assessing the relevance of a given extended collocation for the purposes of exemplification to be included in a dictionary.

3) Most of the included extended collocations will consist of three or four components, i.e. the prototypical (=binary) collocation extended by one or two constituents, whether collocational (lexical), colligational (grammatical), or both. This does <u>not</u> mean that longer sequences, while being uncommon, are automatically out of the question; however, if they are included (on account of being recurrent and salient as units), they are likely to illustrate primarily typical and predictable syntactic strings rather than facts of phraseology, with the  proviso that the dividing line between the two is far from being clear – on the contrary, there is a cline, with the binary collocation and the lengthy extended collocation being merely the two end-points.[18]

---

[17]  As shown e.g. in the standard COCA (450-million-word) corpus of American English (http://corpus.byu.edu/coca/) covering the period of 1990 – 2012.

[18]  That the makeup of such units can be problematic can be seen even in glossary-type analyses of certain terminologies;

To illustrate the above recommendations, here are just two examples (merely a foretaste of the things to come) of the relevant parts of a possible treatment of aspects of extended collocability[19] to be offered within the relevant senses of two specific nominal entry-articles considered in their sporting senses, as drafted primarily for encoding purposes in the general bilingual Slovene-English dictionary:

> **gol**  *sam* [...]  *prejeti/dobiti poceni gol*   *to concede a soft goal*
> *kmalu prejeti/dobiti gol*   *to concede an early goal*
> *doseči gol na gostovanju*   *to score an away goal*
> *razveljaviti prvi gol*   *to disallow the first goal*
> *zaostajati za dva gola*   *to be two goals down*
> *imeti prednost treh golov*   *to be three goals up*

> **vodstvo**  *sam* [...]  *kmalu priti v vodstvo*   *to open/take an early lead*
> *kmalu priti do tesnega vodstva*   *to open up a small lead*
> *priti do suverenega*[20] *vodstva*   *to build up a commanding lead*
> *zapraviti prepričljivo vodstvo*   *to blow a comfortable lead*
> *povišati vodstvo iz prvega polčasa*   *to extend the first-half lead*
> *boriti se za ohranitev vodstva*   *to struggle to stay in the lead.*

# References

Biber, Douglas, Stig Johansson, Geoffrey N. Leech, Susan Conrad, and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow, Essex: Pearson Education / Longman.

Biber, Douglas, and Federica Barbieri. 2007. "Lexical Bundles in University Spoken and Written Registers." *English for Specific Purposes* 26 (3): 263-86.

The Corpus of Contemporary American English (COCA). Available at http://corpus.byu.edu/coca/. Accessed July 28, 2014.

Danielsson, Pernilla. 2007. "What Constitutes a Unit of Analysis in Language?" *Linguistik Online* 31 (2): 17-24. Available at http://www.linguistik-online.de/31_07/index.html. Accessed on July 28, 2014.

Fox, Chris, and Rosalind Combley, eds. 2014. *Longman Dictionary of Contemporary English*. 6th ed. Harlow, Essex: Pearson Education.

Francis, Gill. 1993. "A Corpus-Driven Approach to Grammar." In *Text and Technology: In Honour of John Sinclair*, edited by Mona Baker, Gill Francis, and Elena Tognini-Bonelli, 137-56. Amsterdam: John Benjamins Publishing.

thus in an exploration of English football jargon (Jellis 2006), the author lists a number of extended units (italicized or boldfaced to show their status as units – only italicized here for the sake of simplicity) including *get into the squad, coming back from injury, [not to be] caught in possession, [to be] on a yellow card, [to be] shown the red card, to head for an early bath, [to be] good in the air*, etc.; however, in other cases parts of expressions are, for reasons unknown to me, not given in italics/bold, implying that they are not really parts of the expressions: to keep the ball *in play*, to avoid the *offside trap, to lay* a ball *off* to a *teammate*, to score *against the run of play*, to blow *the final whistle*, the ball slams into *the wall*, to make *a late tackle* in the *penalty area*, to score the *winning goal*, for example.

[19]   Let us merely note here in passing that even at the basic – binary – level of collocability, there are in most of the general-purpose advanced learners' English dictionaries  and in quite a few general-purpose bilingual dictionaries cases of a curious lack of the "reciprocity" logic: Thus e.g. while virtually all such dictionaries routinely list, under *goal*, the collocation *to score a goal*, many of them fail to record the "reciprocal" collocation *to concede a goal*.

[20]   Frequently used these days by some Slovene tennis commentators (on TV), apparently in a variety of senses including 'comfortable', 'substantial', 'obvious'.

Gabrovšek, Dušan. 2011. "Reduced Collocational Identity or Identities in English?" In *Identities in Transition in the English-Speaking World*, edited by Nicoletta Vasta, Anotonella Riem Natale, Maria Bortoluzzi in Deborah Saidero, 105-19. All Series, no. 8. Udine: Forum.

— — —. 2012. "Experiential Phraseology and the Bilingual Dictionary." Paper presented at EUROPHRAS 2012, Maribor, *Phraseology and Culture*, University of Maribor, August 27-30, 2012.

Gledhill, Christopher J. 2000. *Collocations in Science Writing*. Language in Performance #22. Tübingen: Gunter Narr.

Hoey, Michael. 2005. *Lexical Priming: A New Theory of Words and Language*. Abingdon, Oxon: Routledge.

Hunston, Susan. 2004. [2005] "The Corpus, Grammar Patterns, and Lexicography." *Lexicographica* 20: 100-13.

Hunston, Susan, and Gill Francis. 2000. *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English.* Studies in Corpus Linguistics, Vol. 4. Amsterdam: John Benjamins Publishing.

Jellis, Susan. 2006. "Lifting the Silverware: An Exploration of English Football Jargon." *MED Magazine* 37 (April). Available at http://www.macmillandictionaries.com/MED-Magazine/April2006/37-Feature-Football-Jargon.htm. Accessed on August 5, 2014.

Lord, Robert. 1994. *The Words We Use*. London: Kahn & Averill.

Mayor, Michael, ed. 2009. *Longman Dictionary of Contemporary English.* 5th ed. Harlow, Essex: Pearson Education.

— — —, ed. 2013. *Longman Collocations Dictionary and Thesaurus.* Harlow, Essex: Pearson Education.

McIntosh, Colin, ed. 2009. *Oxford Collocations Dictionary for Students of English*. 2nd ed. Oxford: Oxford University Press.

— — —, ed. 2013. *Cambridge Advanced Learner's Dictionary*. 4th ed. Cambridge: Cambridge University Press.

Partington, Alan. 1998. *Patterns and Meanings: Using Corpora for English Language Research and Teaching*. Studies in Corpus Linguistics, Vol. 2. Amsterdam: John Benjamins Publishing.

Pearce, Michael. 2007. *The Routledge Dictionary of English Language Studies*. Abingdon, Oxon: Routledge.

Rundell, Michael, ed. 2007. *Macmillan English Dictionary for Advanced Learners.* 2nd ed. Oxford: Macmillan Education.

— — —, ed. 2010. *Macmillan Collocations Dictionary*. Oxford: Macmillan Education.

Sinclair, John McH. 1991. *Corpus, Concordance, Collocation*. Describing English Language. Oxford: Oxford University Press.

— — — . 1996. "The Search for Units of Meaning." *Textus* 9 (1): 75-106. Reprinted in Sinclair 2004, 24-48.

— — — . 1998. "The Lexical Item." In *Contrastive Lexical Semantics*, edited by Edda Weigand, 1-24. Current Issues in Linguistic Theory, Vol. 171. Amsterdam: John Benjamins Publishing. Reprinted in Sinclair 2004, 131-48.

— — —. 2004. *Trust the Text: Language, Corpus and Discourse*. Edited by Ronald A. Carter. London: Routledge.

— — —, ed. 2001. *Collins COBUILD English Dictionary for Advanced Learners*. Glasgow: HarperCollins Publishers.

Stubbs, Michael. 2002. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell Publishing.

— — —. 2004. "Language Corpora." In *The Handbook of Applied Linguistics*, edited by Alan Davies and Catherine Elder, 106-32. Blackwell Handbooks in Linguistics #17. Malden, MA and Oxford: Blackwell Publishing.

Turnbull, Joanna, and Diana Lea, eds. 2010. *Oxford Advanced Learner's Dictionary of Current English*. [By] A. S. Hornby. 8th ed. Oxford: Oxford University Press.