

JAKUB SLÁMA

Department of Contemporary Lexicology and Lexicography
Czech Language Institute of the Czech Academy of Sciences, Praha
e-mail: slama@ujc.cas.cz
ORCID 0000-0002-6555-0471

THE SKEWED FREQUENCY HYPOTHESIS AND THE IDENTIFICATION OF VALENT NOUNS IN ENGLISH¹

ABSTRACT: I propose the Skewed Frequency Hypothesis, according to which the distribution of corpus types following immediately after the occurrences of a valent noun is significantly skewed, unlike with a valent nouns. The question explored in this paper is whether this observation can be used to retrieve a list of valent nouns from a corpus. Two simple ratios are used to measure the skewness of the distribution of types, and it is shown that the two ratios are correlated and serve as somewhat reliable cues of noun valency. The cases are discussed in which the two ratios suggest conflicting results, and it is argued that these cases serve as one of the arguments for viewing noun valency in English as a scalar phenomenon rather than a binary possibility (valent vs. a valent). Finally, I also discuss some other issues concerning noun valency in English, and based on the data, I suggest treating noun valency in terms of Construction Grammar.

KEYWORDS: argument structure, Construction Grammar, colligation, complements, noun valency

SŁOWA KLUCZOWE: struktura argumentowa, gramatyka konstrukcyjna, koligacja, określenia predykatywne, walencja rzeczowników

¹ The preparation of this article was financed within the statutory activity of the Czech Language Institute of the Czech Academy of Sciences (RVO No. 6837809).

1. Introduction

If an English noun has a valency frame (or, an argument structure), thus requiring an argument, this argument is typically expressed as a complement that takes the form of a prepositional phrase or another unit introduced by a grammatical word. It therefore follows that valent nouns might colligate with certain grammatical words. Based on this, I propose what I refer to as the Skewed Frequency Hypothesis: a valent noun should enter colligational relations, based on which the frequency distribution of the types following the noun in a corpus should be skewed, and the position immediately following the noun should be less variable (type-wise) than after an aivalent noun. I attempt to operationalize the relevant variables and, using data from the British National Corpus² (BNC), to test whether this hypothesis can prove useful for the identification of valent nouns in a corpus.

2. Background

In this section, I first elaborate on the central concept of noun valency (2.1), and after briefly discussing the relevant phenomenon of colligation (2.2), I illustrate what I refer to as the Skewed Frequency Hypothesis, explaining its rationale and its relevance to the procedure of identifying valent nouns in English (2.3).

2.1. Noun valency and complementation

Semantically, two types of nouns can be distinguished: sortal nouns and relational nouns. Sortal nouns simply denote (or classify) entities, e.g. *chair* and *table* (Plag 2003, p. 148). Relational nouns, on the other hand, “denote relations between a specific entity and a second one” (Plag 2003, p. 148). For example, *father* and *surgery* are relational nouns because, respectively, to qualify as a father, one has to be a father of someone (or, metaphorically, of something), and a surgery always has to be performed on something. This necessary “something” is linguistically expressed as a complement (or, to use a different term, an argument) of the relational noun; the terms *complement* and *argument* are often used interchangeably (e.g. Resi 2014, p. 170). Even though both of these terms “can refer to an element in the semantic representation and also to a syntactic entity” (Van Valin, LaPol-la 1997, pp. 27–28) and, being ambiguous, might require further discussion, in this paper I will simply use *argument* to refer to the above-mentioned necessary “something” on the semantic level, and *complement* to refer to the syntactic entity

² The BNC was accessed through the KonText interface of the Czech National Corpus project, available online from <<https://kontext.korpus.cz>>.

referring to the argument. Thus, in *the father of modern linguistics*, *father* is a relational (and hence, in my view, valent) noun, *of modern linguistics* is its complement, and *modern linguistics* refers to the argument of the noun *father*.

The obvious question follows of how to differentiate a complement of a noun from an optional modifier, exemplified respectively in *the destruction of the rainforests* and *rainforest destruction in the Amazon Jungle*. Apart from my intuition (as to what nouns clearly do not have valency), I will rely here on the dialogue test and on the contextual non-boundness test. If one says *My friends have already arrived* and another person asks where the friends have arrived, *I don't know* is not a felicitous answer, and thus the goal is considered as an obligatory argument – this is the point of the dialogue test (Panevová 2014, p. 5). Similarly, based on the dialogue test, one can say that e.g. the noun *attempt* requires an argument:

- (1) Many of these species perished in the attempt. (BNC) – The attempt to do what?
The attempt at what? – *I don't know.

This relates to what I have referred to as the contextual non-boundness test, whose relevance I have illustrated elsewhere (Sláma 2018a; Sláma 2018b). The usual claim is that the complement of a noun is never obligatory (e.g. Lieber 2018, p. 109), but I have shown that when a valent noun is contextually non-bound, its complement can in fact be obligatory. For example, when one examines the instances of *an ability* in the written subcorpus of the BNC in which *an* marks the first use of the noun whereby its referent is introduced into the discourse (thus excluding e.g. instances of the semi-anaphoric *such an ability*), it turns out that every single one of these 230 instances of *an ability* has an overt complement. Therefore, it seems justified to claim that *ability* indeed has valency – contrary to some traditional accounts which attribute noun valency only to complex event nominals, a claim made most influentially by Grimshaw (1990, p. 45).

2.2. Colligation

As Lehecka (2015) points out, the term *colligation* “has been used in a large number of different senses” (also cf. Hoey 2005, p. 43). Nowadays it seems to be the case that when studying colligation, linguists most often focus on textual colligation, i.e. the fact that lexical items “are primed to occur in or avoid, certain positions within the discourse” (Hoey 2005, p. 13). This is the type of colligation that is examined in many corpus studies, including, for instance, the one by Dong and Buckingham (2018). However, I use the term *colligation* in the sense pioneered by Firth (1968, p. 182), who talks of “the interrelation of a set of grammatical categories transcending the actual words which may fall into those categories.” Very generally, this means that, for instance, nouns colligate with adjectives and

verbs tend to colligate with adverbs; but, importantly for this paper, this can also be applied to individual phrases or words, and one can for example state that the noun *window* tends to colligate with prepositions (Baker et al. 2006, p. 36), as witnessed by contexts such as *casual passers-by*, *looking in the window at the menu* (BNC). Since patterns of noun complementation are more or less limited to *to*-infinitives, prepositional phrases, and *that*-clauses (Hunston, Francis 1999, p. 6), it is clear what kinds of words might be expected to be found immediately after valent nouns.

2.3. The Skewed Frequency Hypothesis

Based on the discussion above, I assume that *table* is a sortal noun that lacks valency properties, and that *ability* is a good example of a relational noun that requires an argument. If we use the written subcorpus of BNC to investigate the colligations³ of these two nouns by a rather crude heuristic, i.e. by looking at what the most frequent lemmas are following the nouns (including punctuation), we get the following results:

<i>table</i>		<i>ability</i>	
.	2,835	<i>to</i>	5,371
,	2,354	<i>of</i>	1,132
<i>and</i>	1,427	.	657
<i>be</i>	599	,	579
<i>in</i>	513	<i>and</i>	469
<i>others</i>	12,843	<i>others</i>	1,734
<i>total</i>	20,571	<i>total</i>	9,942

Table 1. The lemmas most frequently following *table* and *ability* in the written BNC

At first blush, there is an obvious difference: while more than half of the instances of *ability* in the written BNC subcorpus are followed by *to* and roughly 65% of the instances of *ability* are followed either by *to* or by *of*, no such dominant pattern can be identified for *table*. The three most frequent lemmas (excluding punctuation) given in the table account for 12.3% of the instances of *table* (including those followed by punctuation), while for *ability* the proportion is 70.1%.

³ By looking at the most frequent lemmas, one seems to focus on collocations rather than colligations; however, as McEnery and Brezina (2019, p. 98) note, “while it is often convenient to think of collocations and colligations as discrete categories, they should more properly be viewed as two poles of a lexicogrammatical continuum, meaning that it is almost inevitable that, in searching for meaning through collocation, we will uncover grammatical information.”

One does not need statistical significance testing to see that the difference can hardly be due to chance.

The question explored in this paper is whether a similar difference can be reliably found across English nouns, and, when turned “inside out”, whether this observation can be used to retrieve a list of valent nouns from a corpus.

3. Data

To collect data, I used the written subcorpus of the BNC, first extracting a list of all nouns from the subcorpus ordered by frequency.

It is a well-documented fact that more frequent words tend to be polysemous (cf. Divjak, Caldwell-Harris 2015, p. 57), i.e. “the different usages or ‘meanings’ of a tool or word increase with frequency” (Zipf 1949, p. 109), and since my approach here is form-based, polysemy might distort the results; therefore, I decided to exclude the 426 nouns from the list that are attested in the subcorpus more than 10,000 times (a rather arbitrary limit) and thus seem most prone to polysemy. Furthermore, the dataset would ideally be of limited size so that it could be examined manually in detail, and so I excluded the nouns from the list that are attested in the subcorpus fewer than 5,000 times (again, this is a rather arbitrary threshold). Finally, I excluded the lemma *no*, abbreviations (e.g. *m.* and *fig.*), and proper nouns, which are “in a way ‘intransitive’ and cannot therefore have any arguments” (Resi 2014, p. 164–165), including e.g. *Germany*, *September*, *America*, *James*, *West*⁴, and *english* (sic).

For each of the 436 remaining lemmas, I extracted a) its frequency, b) the number of various types occurring in the R1 position (i.e. following immediately after the noun), and c) the lemmas and frequencies of the three most frequent grammatical words in the R1 position. I added to the data set three columns: the sum of the frequencies of the three most frequent grammatical words, the ratio of this sum and the total absolute frequency of the lemma, and the ratio of the number of types in the R1 position and the absolute frequency of the lemma. For convenience sake, I will refer to the former ratio as r_1 and to the latter ratio as r_2 . For instance, *ability* is most frequently followed by *to*, *of*, and *and*, as mentioned above, and the instances of *ability to*, *ability of*, and *ability and* account for 70.1% occurrences of *ability*; r_1 for *ability* is thus roughly 0.7. Overall, *ability* is followed by 327 different types, and so with its absolute frequency of 9,942 occurrences, its r_2 is roughly 0.03.

⁴ Arguably, however, *North*, *East*, *South*, and *West*, although lemmatized as proper nouns in the BNC, seem to take arguments; cf. Sláma and Štěpánková (2019) on the valency of the corresponding adverbs in Czech.

The interpretation of r_1 and r_2 is rather straightforward: the higher the r_1 for a noun, the more often is the noun followed by one of the three lemmas occurring most frequently in the R1 position, and the higher the r_2 for a noun, the higher the relative number of different lemmas immediately following the noun.

Finally, I intended to annotate each of the 436 nouns for its type (sortal vs. relational), but this turned out to be rather problematic; while the distinction is typically clear with the usual textbook examples, its application to a sample of nouns seems rather subjective and hence problematic in a plethora of cases. Therefore, I did not include this distinction in the data set.

4. Results

At first blush, the two ratios seem to be somewhat suggestive of whether the noun requires a complement (and thus has an argument structure). For instance, while r_1 is highest for *lack* (0.98), *variety* (0.76), *proportion* (0.73), *absence* (0.71), and *combination* (0.70), which could be considered (mostly not controversially) as valent, the ratio is very low for nouns such as *sir* (0.004), *football* (0.091), *telephone* (0.093), *metal* (0.102), and *cash* (0.118), which can be described as not requiring a complement. Similarly, r_2 is lowest for arguably valent nouns such as *lack* (0.01), *importance* (0.03), and *ability* (0.03), and highest for clearly avalent nouns such as *metal* (0.25), *adult* (0.16), and *gold* (0.16).

Furthermore, it is not surprising (yet it is not self-evident either) that there is a negative correlation⁵ between the two ratios⁶ ($r(434) = -0.6768693$; $p < 2.2e-16$, the 95% confidence interval of correlation is $[-0.7247141, -0.6225289]$), summarized in the following scatterplot (Figure 1), where each data point corresponds to one noun. Since the two ratios are correlated, they should serve as comparably reliable indicators of whether a noun has an argument structure. In the following two sections, I will thus focus on the reliability of the ratios and on the cases in which the ratios suggest contradictory results (i.e. one suggests that the noun might be valent and the other suggests the opposite).

⁵ To analyze the data, I used R, a free software environment for statistical computing (R Development Core Team 2008).

⁶ The values of both ratios are normally distributed – with the sample size of $N = 436$, this is not necessary to test, but still, the Shapiro-Wilk normality test confirms that the values are indeed normally distributed (r_1 : $W = 0.8745$, $p < 2.2e-16$; r_2 : $W = 0.98029$, $p = 1.199e-05$).

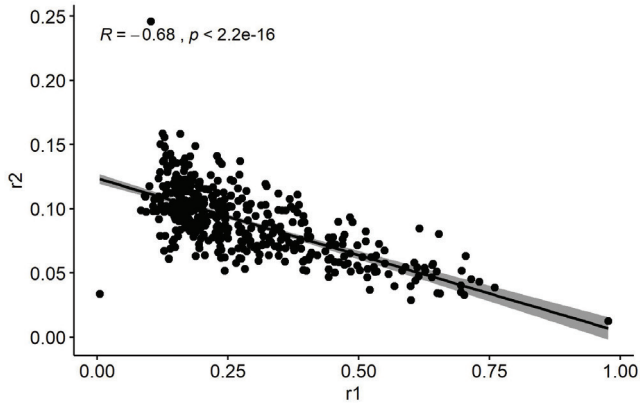


Figure 1. The correlation between r_1 and r_2

4.1. Reliability of r_1

As stated above, higher values of r_1 are to be found with nouns that are typically followed by a grammatical word (typically a preposition) introducing a complement. The thirty nouns with the highest r_1 include:

noun	r_1	noun	r_1	noun	r_1
<i>lack</i>	0.977	<i>introduction</i>	0.653	<i>means</i>	0.612
<i>variety</i>	0.760	<i>possibility</i>	0.650	<i>intention</i>	0.608
<i>proportion</i>	0.731	<i>understanding</i>	0.649	<i>copy</i>	0.606
<i>absence</i>	0.715	<i>access</i>	0.642	<i>importance</i>	0.600
<i>combination</i>	0.704	<i>bit</i>	0.640	<i>emphasis</i>	0.599
<i>ability</i>	0.701	<i>implication</i>	0.640	<i>sign</i>	0.589
<i>reduction</i>	0.696	<i>majority</i>	0.627	<i>desire</i>	0.582
<i>beginning</i>	0.695	<i>impact</i>	0.625	<i>consequence</i>	0.568
<i>rise</i>	0.695	<i>pair</i>	0.616	<i>commitment</i>	0.557
<i>aim</i>	0.655	<i>presence</i>	0.615	<i>concept</i>	0.550

Table 2. The thirty nouns with the highest values of r_1

If one applies the tests mentioned above, all the nouns seem to qualify as valent. For instance, there cannot be a lack unless it is the lack of something, and this is reflected in the fact that from the 1,104 instances of *a lack* in the written BNC, 1,089 are immediately followed by *of*, two are false positives (*A lack lustre scheme; If A lacks the legal capacity...*), and the remaining examples either do not feature

a lack as new in the discourse (cf. example 2, in which the semi-anaphoric *such a lack* is actually preceded by another instance of *lack*, from which it is clear the lack whereof is at stake), or feature a generalized use of *lack* (3), comparable to the intransitive use of verbs with indefinite null complements such as *eat* – and just like the intransitive use of *eat* does not disqualify the verb from having an argument structure, we should acknowledge that examples like (3) do not disqualify nouns like *lack* from having one:

- (2) Sweeney shocks because he shows so clearly the effect of **lack of themis**. Harrison had written of **such a lack** among the Cyclopes. (BNC)
- (3) It acts as a supplement to **a lack**, something missing, on the part of the signified thing. (BNC)

Therefore, based on the dialogue test and on the related test concerned with context boundness, discussed above, *lack* can be seen as valent – and the same applies to all the other nouns in Table 2. However, what is rather interesting is the fact that it applies to the nouns in varying degrees. While *lack* and *ability*, when preceded by the indefinite article marking contextual non-boundness, seem to require an overt complement, with a noun like *copy* the situation is not as straightforward. The phrase *a copy* is followed by *of* only in 1,589 (74%) of its 2,136 occurrences in the written BNC subcorpus, and while in the remaining examples a complement can often be found (4), there are still instances when there simply is not one, as in (5):

- (4) It is the one **of which** a copy is used today. (BNC)
- (5) The completed NEWOED2 form should be returned to the New OED Computer Group representative making sure **a copy** is retained. (BNC)

In (5), it is clear that *a copy* refers to a copy of the aforementioned form, but the noun is still used with the indefinite article and without an overt complement. Based on the dialogue test, however, we can maintain that *copy* requires an argument and is, indeed, valent.

Table 3 gives the thirty nouns for which r_1 reaches the lowest values:

noun	r_1	noun	r_1	noun	r_1
<i>gold</i>	0.1284	<i>drug</i>	0.1226	<i>speaker</i>	0.1086
<i>crown</i>	0.1282	<i>media</i>	0.1209	<i>phone</i>	0.1071
<i>screen</i>	0.1280	<i>prison</i>	0.1205	<i>afternoon</i>	0.1067
<i>winter</i>	0.1279	<i>nation</i>	0.1204	<i>metal</i>	0.1024
<i>pension</i>	0.1278	<i>blood</i>	0.1203	<i>railway</i>	0.0999
<i>silence</i>	0.1277	<i>radio</i>	0.1200	<i>telephone</i>	0.0936

noun	r_1	noun	r_1	noun	r_1
<i>traffic</i>	0.1259	<i>cash</i>	0.1177	<i>consumer</i>	0.0920
<i>weekend</i>	0.1259	<i>league</i>	0.1175	<i>football</i>	0.0905
<i>acid</i>	0.1257	<i>cabinet</i>	0.1170	<i>insurance</i>	0.0838
<i>adult</i>	0.1251	<i>while</i>	0.1136	<i>sir</i>	0.0046

Table 3. The thirty nouns with the lowest values of r_1

While most of the nouns are rather clearly avalent sortal nouns, there are some less clear cases. For instance, *crown* can be found in phrases such as *a crown of thorns*, *a crown of wild flowers*, and the like, in which the status of the *of*-phrase might be somewhat disputable (but it seems appropriate to conclude that the phrase is a modifier, not a complement, based on the dialogue test, among other things). The noun *winter* is perhaps best seen as avalent; however, in some literal as well as metaphoric contexts in which *winter* receives a meronymic (i.e. part-whole) interpretation, it does take what seems to be a complement:

- (6) The underlying cause for this decision was the awful damage caused by the savage **winter** of 1709. (BNC)
- (7) When a man reaches the **winter** of his life, there's nothin' he can look forward to but death. (BNC)

Another problem might be presented by the noun *silence*, which seems to be mostly avalent, but in some contexts takes a complement:

- (8) Lord Donaldson said that he could now break the traditional **silence** of judges about their colleagues... (BNC)

The noun *weekend* is also noteworthy: while it is typically used without a complement, in some contexts it seems to parallel some valent nouns, referring to containers (9), including e.g. *a cup of tea*, and some meronyms (10), i.e. nouns referring to parts of something, that can be seen as valent, including *part* itself but also *winter* in examples 6 and 7:

- (9) a weekend of celebration (BNC)
- (10) the weekend of the first week of September (www.dailystar.co.uk)⁷

Finally, *blood*, *speaker*, and *consumer* can be used with *of*-phrases which seem to possibly qualify as complements (even though in (11), the *of*-phrase might be seen as a modifier, and according to the dialogue test, it does not seem to be obligatory):

- (11) they wore robes stained scarlet from the **blood** of their victims (BNC)

⁷ Retrieved from Fulda 2020.

- (12) native **speakers** of English (BNC)
 (13) the greatest **consumers** of wheat fibre are the Slavs (from Yugoslavia) (BNC)

Scrutiny of the complete dataset suggests that r_1 is rather reliable in that the nouns with r_1 of 0.25 or higher⁸ can be (mostly not controversially) considered as valent. However, the discussion above illustrates that the reliability of the ratio is somewhat limited in that a lower value of the ratio does not guarantee that the noun cannot take a complement.

4.2. Reliability of r_2

As stated above, the two ratios are correlated; however, there are some notable discrepancies between them. For instance, the avalent noun *sir* with a very low r_1 has the fourth lowest r_2 , following only after three clearly valent nouns, *lack*, *importance*, and *ability*. Similarly, while e.g. *discipline*, *past*, *peace*, *silence*, *public*, *accident*, and *smile* are expected to be avalent based on their r_1 values, they rank among clearly valent nouns according to r_2 . While these nouns (for a brief discussion of *silence*, see the previous section) seem to be mostly avalent, in some contexts some of them appear to take a complement (note, however, the coordination of *training* and *discipline* in (14)):

- (14) the poor training and **discipline** of the security forces (BNC)
 (15) a dark stain on the **past** of a very important German industrialist (BNC)
 (16) the **peace** between Greeks was breaking down too (BNC)
 (17) brief knowing **smiles** of senior academics on committees (BNC)

On the other hand, some nouns, including *location*, *victim*, *transfer*, *guide* and *publication*, would be expected to be valent based on their r_1 values (and also based on the dialogue test) but rank as avalent if r_2 is used.

In conclusion, while r_1 and r_2 correlate and seem to be relatively useful indicators of argument structure when trying to extract a list of valent nouns from a corpus, their reliability is far from absolute. Interestingly, when the two ratios are in conflict, r_1 of 0.25 or more is generally a good marker that the noun is likely to be valent, but when r_1 is lower and the noun seems not to have valency, although r_2 suggests that it could be expected to take a complement, the noun generally seems not to take a complement, but occasionally does.

⁸ Nouns with r_1 just above this threshold include *distance* (0.251), *surface* (0.255), *reading* (0.256), *base* (0.267), and *author* (0.269). Nouns with r_1 just below this threshold include *culture* (0.244), *truth* (0.243), *facility* (0.243), *crowd* (0.241), and *pain* (0.238). Naturally, however, the value of 0.25 is not a strict threshold dividing all valent nouns from all avalent nouns.

5. Discussion

What I see as the most important point implied in this paper is the observation that it might be more adequate to treat noun valency as a scalar phenomenon, rather than a binary one. That is, while the traditional approach consists in stating that a noun does (or does not) have valency properties and thus takes (or does not take) a complement, a different view might be somewhat more appropriate. It seems to be clear that some nouns, like *ability*, do have an argument structure, while others, like *coffee*, do not⁹. However, with slight semantic changes, the valency properties might be affected: when *ability* is used to mean ‘ability in general, ability at anything one can have ability at’, typically in texts on education, as in (18), it does not seem to require an argument (just like *lack* in example 3 above):

- (18) Thus an **ability** represents what a person can do now, whereas a capacity is essentially a potential. (BNC)

Furthermore, there seem to be a host of nouns which sometimes seem to require an argument (or, less frequently, arguments), but typically do not – often without obvious semantic changes. For example, there are 865 instances of the form *selfie* in the Corpus of Contemporary American English¹⁰ (COCA), and some of the 29 of those of them followed by *of* seem to be very good examples of *selfie* followed by a complement, e.g.:

- (19) Luliana had been contractually obligated to post a **selfie** of her eating a slice of pink cake with glitter frosting (COCA)
- (20) a **selfie** of the two of them in some restaurant (COCA)
- (21) A brave model from London managed to save herself from a life-threatening hostage situation after secretly sending a **selfie** of her bruised, battered face to a friend. (COCA)

This might be taken to mean that *selfie* requires an argument, perhaps on analogy with nouns such as *photograph*; after all, the dialogue test tells us that this is the case (*Can we take a selfie? – Of whom? – *I don’t know.*). However, I think it is rather questionable whether it is useful to say that *selfie* has an argument structure when it is in fact dispreferred to use an overt complement (apart from the numbers given above, it also seems rather unusual to say something like *I will take a selfie of myself*), and in an informal poll I drew among ten linguists, seven answered that they would not see *selfie* as a valent noun in English. It is not clear whether it is more adequate to treat *selfie* as a valent, admitting that it still

⁹ In phrases like *a coffee of the iced variety* (taken from the Corpus of Contemporary American English (Davies 2008)), the *of*-phrase is clearly a modifier, not a complement.

¹⁰ Available online from <<https://www.english-corpora.org/coca/>>.

occasionally takes what seems to be a complement (and that the dialogue test might not be exactly helpful here), or to treat *selfie* as valent, while acknowledging that this is somewhat futile, given the fact that the noun typically does not, or possibly even cannot, take an overt complement.

To give another example, it is quite common for certain nouns referring to entities that can be conceptualized (and used) as containers to vacillate between being a-valent (22) and taking an argument (23):

(22) you broke her **mug** (COCA)

(23) He'd return with a **mug** of tea. (COCA)

A noun like *backpack* is not typically followed by a complement (only 39 of the 6,324 instances of the form *backpack* are immediately followed by *of* in COCA), but when it is, the *of*-phrase can be quite naturally interpreted as a complement on a par with the complements of other container nouns:

(24) I know this individual had a **backpack** of explosives. (COCA)

Arguably, this is not enough for *backpack* to be seen as polysemous, and so the traditional assumption that a change in valency equals a change in meaning is not particularly helpful here – in part because such an approach is rather circular, claiming that the predicate has multiple valency frames because it has multiple meanings, and at the same time deriving the claim that the predicate has multiple meanings from the observation that it has multiple valency frames; cf. Goldberg (1995, p. 11).

Generally, having valency is usually equated with taking a complement (or complements) referring to an argument (or arguments). Hellan et al. (2017, p. 1) state explicitly that “the notion ‘argument’, traditionally used in the logical disciplines, has become equivalent to ‘valency-bound’.” To illustrate this, it can be mentioned that, for instance, Klímová et al. (2016, p. 1) define noun valency as “the number, type and form of arguments that are bound to a noun.” The equation of having argument structure and requiring an argument is implied in the definitions of valency in general, cf. e.g. Haspelmath and Hartmann's (2015, p. 42) delineation: “The valency of a verb is the range of syntactic properties of other elements of the clause that depend on the particular choice of verb, i.e. that are verb-specific. These other elements of the clause are called arguments.”

However, some nouns (like *backpack*) seem not to be valent, but still take an argument, without necessarily being polysemous. I believe that the problem has a rather obvious solution: we can admit that a sortal noun, albeit not inherently valent, can enter a syntactic construction on analogy with semantically similar nouns that are often used in the construction, and the construction can

contribute an argument. This is in fact rather common with verbs, as illustrated in the large body of research conducted within Construction Grammar, with Goldberg (1995) as one of the pioneering works in this regard. For instance, the verbs *sneeze*, *smile*, and *laugh* are typically intransitive, but they may enter various argument structure constructions, in which they appear to have a new valency without any change in lexical meaning (the examples are cited after Goldberg (2006, p. 6)):

(25) He **sneezed** his tooth right across town.

(26) She **smiled** herself an upgrade.

(27) We **laughed** our conversation to an end.

Again, if one claims that e.g. *sneeze* in (25) has a new sense ('to cause something to move by sneezing'), one commits the circularity fallacy mentioned above, proposing this new sense only on the basis of the unusual valency frame, but also using this presumably new sense to explain why the verb occurs in the unusual valency frame.

Goldberg and others use examples like (25), (26), and (27) as one of the arguments for a constructionist view of argument structure rather than the lexical(ist) view¹¹ (which is based on the notion that the valency of a lexical unit is determined by the lexical unit alone). For related reasons, Haugen (2013) argues that valency of adjectives supports a constructional approach to valency, and I believe that the points raised above might present a good argument for a constructional approach to noun valency too.

A noun like *backpack* is primarily sortal and avalent, but when it enters the construction with the container interpretation, the construction contributes what appears to be a complement of the noun. Since the noun does not have the container interpretation unless it is used in the construction, what might be perceived as a new meaning is more adequately seen as a property of the whole construction, not of the lexical item alone. This description avoids what has been called the polysemy fallacy, i.e. viewing contextually-bound uses of a lexical item as instances of polysemy (Haugen 2013, p. 63). On the constructional view, the noun has only one meaning, but it can be used in various constructions, which can lead to slightly different semantic interpretations of the same sense of the noun.

To summarize the discussion, noun valency in English appears to be more adequately treated as a scalar phenomenon (rather than a yes-no phenomenon), and nouns that appear to be inherently avalent sometimes seem to be able to

¹¹ "The basic idea behind lexical models of valency is that the predicator, the valency carrier, determines how it is to be complemented. However, individual predicators are typically found in several valency patterns, and this is also valid for adjectival predicators." (Haugen 2013, p. 47)

take arguments contributed by a construction possibly associated with a specific interpretation (e.g. the container interpretation or the meronymic interpretation), but not necessarily with a new meaning. I have also suggested that in view of the data presented above, noun valency (in English) might be described more adequately within the constructionist framework.

Sources

- The British National Corpus*, version 2 (BNC World). Distributed by Oxford University Computing Services on behalf of the BNC Consortium. Institute of the Czech National Corpus, Prague 2001. Available online from <<http://www.korpus.cz>>. Last accessed August 15, 2020. (BNC)
- Davies, M. (2008). *Corpus of Contemporary American English*. Available online from <<https://www.english-corpora.org/coca/>>. Last accessed August 15, 2020. (COCA)
- Fulda, O. (2020). Premier League 2020/21 season ‘must start by first week of September’. *Daily Star* online. Retrieved from <<https://www.dailystar.co.uk/sport/football/premier-league-202021-season-must-21887626>>. Accessed August 3, 2020.

Bibliography

- Baker, P. et al. (2006). *A Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Divjak, D., Caldwell-Harris, C. (2015). Frequency and Entrenchment. In: E. Dąbrowska & D. Divjak (eds.), *Handbook of Cognitive Linguistics* (53–74). De Gruyter Mouton.
- Dong, J., Buckingham, L. (2018). The textual colligation of stance phraseology in cross-disciplinary academic discourse, *International Journal of Corpus Linguistics*, 23(4), 408–436.
- Firth, J.R. (1968). A synopsis of linguistic theory, 1930–1955. In: F. Palmer (ed.), *Selected Papers of J. R. Firth 1952–1959* (168–205). London: Longmans.
- Goldberg, A.E. (1995). *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago: The University of Chicago Press.
- Goldberg, A.E. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Grimshaw, J. (1990). *Argument Structure*. Cambridge: MIT Press.
- Haspelmath, M., Hartmann, I. (2015). Comparing verbal valency across languages. In: A.L. Malchukov. B. Comrie (eds.), *Valency Classes in the World's Languages* (41–71). Berlin: De Gruyter Mouton.
- Haugen, T.A. (2013). Adjectival valency as valency constructions: Evidence from Norwegian, *Constructions and Frames*, 5(1), 35–68.

- Hellan, L. et al. (2017). Issues in contrastive valency studies. In: Hellan, L. et al. (eds.), *Contrastive Studies in Verbal Valency* (1–24). Amsterdam: John Benjamins Publishing Company.
- Hoey, M. (2005). *Lexical Priming: A New Theory of Words and Language*. London: Routledge.
- Hunston, S., Francis, G. (1999). *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins.
- Klímová, J. et al. (2016). Towards a Corpus-based Valency Lexicon of Czech Nouns. In: *GLOBALEX 2016: Lexicographic Resources for Human Language Technology. GLOBALEX Workshop 2016* (1–7).
- Lehecka, T. (2015). Collocation and colligation. *Handbook of Pragmatics Online*. <<https://benjamins.com/online/hop/articles/col2>>.
- Lieber, R. (2018). *English Nouns. The Ecology of Nominalization*. Oxford: Oxford University Press.
- McEnergy, T., Brezina, V. (2019). Collocations and colligations: Visualizing lexicogrammar. In: B. Busse, R. Moehlig-Falke (eds.), *Patterns in Language and Linguistics* (97–124). Berlin–Boston: De Gruyter Mouton.
- Panevová, J. (2014). Contribution of valency to the analysis of language. In: O. Spevak (ed.), *Noun Valency* (1–17). Amsterdam–Philadelphia: John Benjamins Publishing Company.
- Plag, I. (2003). *Word-Formation in English*. Cambridge–New York: Cambridge University Press.
- R Development Core Team (2008). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Austria: Vienna.
- Resi, R. (2014). Noun phrasal complements vs. adjuncts. In: O. Spevak (ed.), *Noun Valency* (161–182). Amsterdam–Philadelphia: John Benjamins Publishing Company.
- Sláma, J. (2018a). *The prepositional phrase with the preposition as a valency complement of nouns*. MA thesis. Prague: Faculty of Arts.
- Sláma, J. (2018b). K (ne)obligatornosti valenčního komplementu anglických substantiv. In: *Kulturní a jazykový prostor v proměnách 21. století. Sborník příspěvků z konference ProfiLingua 2018* (129–142). Pilsen: University of West Bohemia.
- Sláma, J., Štěpánková, B. (2019). On the Valency of Various Types of Adverbs and Its Lexicographic Description, *Jazykovedný časopis*, 70(2), 158–169.
- Van Valin, R.D., LaPolla, R.J. (1997). *Syntax, Structure, Meaning and Function*. Cambridge: Cambridge University Press.
- Zipf, G.K. (1949). *Human Behavior and the Principle of Least Effort*. Cambridge, Mass.: Addison-Wesley Press.

SKEWED FREQUENCY HYPOTHESIS I IDENTYFIKACJA RZECZOWNIKÓW WALENCYJNYCH W JĘZYKU ANGIELSKIM

ABSTRAKT: W niniejszym artykule przedstawiam *skewed frequency hypothesis*, według której dystrybucja typów korpusowych następujących bezpośrednio po wystąpieniu rzeczownika walencyjnego jest znacząco skośna, w przeciwieństwie do rzeczowników niewalencyjnych. Kwestią rozpatrywaną w niniejszym artykule jest ukazanie, czy ta obserwacja może być zastosowana do wyszukiwania w korpusie listy rzeczowników walencyjnych. W celu zmierzenia skośności rozkładu typów zostały zastosowane dwa proste wskaźniki i wykazano, że owe dwa wskaźniki są skorelowane oraz stanowią w miarę niezawodne wskazówki walencji rzeczowników. Omówiono przypadki, w których dwa wskaźniki wskazują sprzeczne wyniki, i stwierdzono, że owe przypadki stanowią jeden z argumentów przemawiających za tym, aby uznać rzeczowniki walencyjne w języku angielskim za zjawisko skalarne zamiast możliwość binarną (walencyjny vs niewalencyjny). Na koniec omówiam także kolejne kwestie dotyczące walencji rzeczowników w języku angielskim i na podstawie danych proponuję traktowanie walencji rzeczowników z punktu widzenia gramatyki konstrukcyjnej.