Attack of the snowclones: A corpus-based analysis of extravagant formulaic patterns

Stefan Hartmann, Tobias Ungerer

Attack of the snowclones: A corpus-based analysis of extravagant formulaic patterns

STEFAN HARTMANN

University of Düsseldorf

TOBIAS UNGERER

Concordia University

The concept of 'snowclones' has gained interest in recent research on linguistic creativity and in studies of extravagance and expressiveness in language. However, no clear criteria for identifying snowclones have yet been established, and detailed corpus-based investigations of the phenomenon are still lacking. This paper addresses this research gap in a twofold way. On the one hand, we develop an operational definition of snowclones, arguing that three criteria are decisive: (i) the existence of a lexically fixed source construction; (ii) partial productivity; (iii) 'extravagant' formal and/or functional characteristics. On the other hand, we offer an empirical investigation of two patterns that have often been mentioned as examples of snowclones in the previous literature, namely [*the mother of all* X] and [X *BE the new* Y]. We use collostructional analysis and distributional semantics to explore the partial productivity of both patterns' slot fillers. In sum, we argue that the concept of snowclones, if properly defined, can contribute substantially to our understanding of creative language use, especially regarding the question of how social, cultural, and interpersonal factors influence the choice of more or less salient linguistic constructions.

KEYWORDS: Construction Grammar, creativity, extravagance, phraseology, snowclones

1. INTRODUCTION

The concept of 'snowclones' has received increased attention in recent constructionist approaches to language (e.g. Traugott & Trousdale 2013, Traugott 2014, Bergs 2018, 2019). Traugott & Trousdale (2013: 150) define snowclones as 'schemas that grow from relatively fixed micro-constructions that are usually formulae or clichés'. Well-known examples include [X *BE the new* Y] and [*the mother of all* X]. While snowclones have become a relatively popular concept in research on the emergence of linguistic constructions (constructionalisation) and in studies of linguistic creativity, the hedges that feature in Traugott & Trousdale's definition (*relatively fixed; usually*) indicate that the term is used rather vaguely in the current literature. To our knowledge, no clear definition offering reliable criteria

599

for identifying snowclones has been put forward so far, which also raises the question of whether the concept is a useful one in the first place. The main goal of this paper is to refine the concept of snowclones by offering a critical review of previous approaches, on the one hand, and a detailed analysis of two particularly frequent snowclones, viz. [*the mother of all* X] and [X *BE the new* Y], on the other. Analysing these constructions can help us understand whether and how snowclones differ systematically from other constructional patterns.

Snowclones have also been discussed in the context of linguistic extravagance, which is another topic that has been gaining popularity in the field (see e.g. Petré 2017, De Wit et al. 2020, Ungerer & Hartmann 2020, Eitelmann & Haumann 2022). Snowclones are usually perceived as constructions that 'stand out' in some way. Understanding snowclones thus also entails understanding the sources from which the patterns derive their 'extravagant' effects. In the remainder of this paper, we will first provide a brief overview of previous research on snowclones (Section 2). Based on the extant literature, we will then propose three criteria for defining the concept (Section 3). In Section 4, we present our corpus-based case studies, focusing particularly on the semantic profiles of the patterns in question. Section 5 brings together the theoretical and empirical findings; moreover, we discuss how snowclones can be distinguished from other types of idiomatic constructions. Finally, we summarise our claims in Section 6 and point out some open questions that could be addressed by future research.

2. SNOWCLONES: A BRIEF HISTORY

The term 'snowclone' was coined in the early 2000s in a 'naming contest' initiated by Geoff Pullum on the linguistics blog *Language Log* (Pullum 2004). Pullum prompted the community to come up with a suitable label for 'a multi-use, customizable, instantly recognizable, time-worn, quoted or misquoted phrase or sentence that can be used in an entirely open array of different jokey variants by lazy journalists and writers' (Pullum 2003b). The term 'snowclone' was suggested by Glen Whitman with reference to Pullum's original example, the journalistic trope that 'if Eskimos have N words for snow, X surely have Y words for Z', such as *If Eskimos have dozens of words for snow, Germans have as many for bureaucracy* (quoted in Pullum 2003a). What makes this pattern a snowclone is the fact that it contains several open slots that can be instantiated by varying lexical fillers.

Since then, more than a hundred *Language Log* posts have been devoted to identifying other snowclones. Moreover, some online collections of snowclones have been created (e.g. O'Connor 2007, Dammerer 2007, and the language-specific Wikipedia entries for 'snowclone'), which comprise several dozens of snowclones from different languages, especially English, German, and French. What these early discussions – as well as more recent case studies such as Tizón Couto's (2021) analysis of COVID snowclones – have in common is that they consider snowclones on a case-by-case basis, illustrating each pattern with a range of examples collected via informal web searches, but mostly focusing on their suspected cultural and

etymological origins. Little attention has been given to common definitional characteristics that would set snowclones apart from other, potentially related types of constructions discussed in the phraseological and constructionist literature, such as proverbs and partially filled constructions like [*What's* X *doing* Y*?*] (Kay & Fillmore 1999; see Section 5 for discussion). Moreover, broader questions about the theoretical status of snowclones as well as their diachronic development are usually backgrounded. A notable exception is Zwicky's (2006b) blog post, in which he proposes a four-step process of how snowclones emerge. According to his account, the creation of snowclones involves (i) a pre-formula stage during which an idea is expressed in various alternative ways (e.g. 'what one person likes, another person detests'); (ii) a first 'fixing' stage during which the idea is captured by a memorable, lexically fixed phrase (e.g. *One man's meat is another man's poison*); (iii) a third stage in which the pattern is extended by the insertion of open slots or playful allusions to it (e.g. *One man's Mede is another man's Persian*); and (iv) a second 'fixing' stage in which the variants become (relatively) routinised into a partially fixed schema with open slots (e.g. [*One man's* X *is another man's* Y]).

More recently, Dancygier & Vandelanotte (2017b) have called for a more systematic, data-driven inquiry into the theoretical status and empirical scope of snowclones. They argue that snowclones should be treated as a distinct construction type with specific formal, semantic, and multimodal features within the framework of Construction Grammar, rather than leaving them 'filed, unanalysed, under "lazy journalists' clichés"'. This, they propose, would allow researchers to 'tap into a quickly growing resource of usage whose popularity calls for a linguistic explanation, relying on various analytical tools (qualitative and quantitative alike)'. In a similar vein, Traugott & Trousdale (2013; see also Traugott 2014, Traugott & Trousdale 2014) regard snowclones as a distinct class of idiomatic constructions and discuss their emergence as a paradigm case of lexical, or 'contentful', constructionalisation. In contrast to traditional accounts of lexicalisation as morphophonological reduction of lexical forms, they argue that contentful constructionalisation involves the creation of partial schemas through processes of both expansion and reduction. Snowclones, under this view, illustrate particularly clearly the process of schema expansion as they arise from inserting variable lexical slots into previously fixed patterns.

Other aspects that have sparked recent research interest are the ways in which snowclones illustrate the interplay of linguistic flexibility and rigidity for the purposes of expressing 'old concepts anew' (Hill 2018) and their role as markers of linguistic creativity (Bergs 2018, 2019). Bergs (2019), for instance, discusses snowclones in the context of Sampson's (2016) distinction between 'F-creativity' (for 'fixed' creativity) and 'E-creativity' (for 'enlarging/extending' creativity). While the former denotes productive extensions of already existing patterns, the latter describes acts of 'rule-breaking' that create novel, unpredictable linguistic material. Bergs suggests that the two types of creativity in fact form a continuum, with snowclones being situated somewhat closer to the F-creative end of the scale. The notion that snowclones are salient and creative constructions will be addressed

601

further in Section 3, where we discuss the 'extravagant' nature of the two constructions we investigate in more detail.

While the above approaches have contributed to a more theoretically informed investigation of snowclones, they are still limited by (i) the lack of a comprehensive definition of snowclones that provides an overview of their prototypical characteristics and (ii) the lack of quantitative corpus-based evidence in support of the theoretical claims, especially regarding the productivity and semantic flexibility of specific snowclones. We address the first of these issues in the following section by developing a criteria-based definition of snowclones. Following that, our two case studies in Section 4 will provide detailed corpus-based evidence about the productivity and semantic profile of two frequent snowclones.

3. DEFINING SNOWCLONES

While several definitions of 'snowclones' have been proposed in previous work, no attempt has yet been made to merge the suggested criteria into a comprehensive and operationalisable account. We will therefore examine a number of previous proposals and extract their common characteristics. By combining them and adding some further modifications, we will arrive at a definition of snowclones that is based on three prototypical criteria. To start with, compare the four previously suggested definitions of snowclones in (1):

(1) (a) 'crisp phrasal templates, with one or two open slots to be filled, generalizing a well-documented specific quotation' (Liberman 2006)
  (b) 'schemas that grow from relatively fixed micro-constructions that are usually formulae or clichés' (Traugott & Trousdale 2013: 150)
  (c) 'patterned figures of speech that empower swapping out words, phrases, or images for one another without breaking the original pattern' (Hill 2018: 423)
  (d) 'a cliché pattern or frame that gives rise to many slightly different variations of this pattern' (Bergs 2019: 176-177)

One common feature pointed out by these statements is that snowclones have a 'source' pattern that is usually lexically fixed. This characteristic functions as the first of the three criteria captured by our definition below. As illustrated by the examples in Table 1, the source instances are typically quotations from real-life historical figures, memorable lines from cultural products (books, songs, films, video games, etc.), or slogans from commercial advertisements. As such, snowclones show a certain degree of intertextuality. They can evoke specific contexts to which the source pattern is tied – for instance, *in law no one can hear you scream* (ENCOW, 118940990aa9a7adee78c6a7be62200ead27) evokes the source *in space no one can hear you scream*, the tagline of the 1979 sci-fi horror film *Alien*.[1]

---

[1] See https://en.wikipedia.org/wiki/Alien_(film)#Release, last checked March 13, 2023.

Snowclone	(Suspected) Lexically Fixed Source
[*the mother of all* X]	*the mother of all battles* (1991 speech by Saddam Hussein)
[X *are from Mars,* Y *are from Venus*]	*Men are from Mars, women are from Venus* (1992 book by John Gray)
[*I am* X*, hear me* Y]	*I am woman, hear me roar* (1971 song by Helen Reddy)
[*One does not simply* VP]	*One does not simply walk into Mordor* (2001 film 'The Lord of the Rings: The Fellowship of the Ring')
[*I'm not an* X*, but I play one on TV*]	*I'm not a doctor, but I play one on TV* (1986 advertisement for Vick's Formula 44 cough syrup)

Table 1
Examples of snowclones and their (suspected) source constructions.

Thus, snowclones can evoke what Schmid (2020) calls pragmatic associations that pertain to contextual factors, answering questions like 'who, where, when, what objects and actions were involved?' (Schmid 2020: 212).[2] Pragmatic associations are of course relevant for all kinds of constructions, but due to their intertextuality, snowclones are more clearly associated with specific contexts than other, less 'specialised', constructions.

The source constructions need to be culturally salient enough to be recognisable by the speech community in which the snowclone is propagated. This does not mean that the source construction necessarily coincides with the earliest attestation of the pattern. For example, [*the mother of all* X] is usually attributed to a quote by Saddam Hussein from 1991, even though some earlier (but culturally less salient) instances of the pattern can be found in corpora (see Section 4.2). Similarly, source constructions can be effective even if their origins are vague or disputed. For instance, [X *BE the new* Y] can be traced back to popular fashion slogans since the 1970s, irrespective of what exact form these slogans took (e.g. *pink is the new black*, or a combination of other colours; see Section 4.3). For the emergence of new snowclones, language users' beliefs about the alleged source of the patterns are of greater relevance than their actual etymological origin.

Moving on to the second defining feature conveyed by all of the above definitions, snowclones are characterised by the extension of the source construction to new instances via partial lexical substitution. Snowclones can thus be regarded as semi-schematic constructions composed of both fixed elements (e.g. *the mother of all*) and open slots (represented by variables such as X and Y). Snowclones can contain one or several variables (see [X *BE the new* Y] with two open slots), which can be either filled by single lexemes or entire phrases (e.g. [*One does not simply* VP] in Table 1). We will argue in Section 4 that analysing these open slots is key to understanding the constructional semantics of a given snowclone.

A third and final feature of snowclones is less explicit in the definitions in (1), but surfaces in expressions like '*crisp* phrasal templates' (our highlighting), 'figures of

---

[2] Thanks to an anonymous reviewer for pointing this out.

603

speech', and, arguably, 'clichés'. The intuition behind these terms is that snow-clones do not simply employ everyday inconspicuous language to communicate an idea, but that they use particularly effective and stylistically striking tools to do so (compare Zwicky 2006b, who refers to snowclones as 'especially apt way[s] of expressing [an] idea'). The concept of 'clichés' is related to this notion of rhetorical effectiveness because clichés are usually regarded as fixed idiomatic expressions whose 'stylistic force' (Howarth 1996: 13) leads to their frequent reuse (or overuse). Nevertheless, we would like to avoid the term here due to its definitional vagueness and potential pejorative connotations.

Instead, we suggest that the striking linguistic features of snowclones can be captured by the concept of 'extravagance'. This term was coined by Haspelmath (1999: 1057) to describe speakers' use of 'imaginative and vivid' language 'in order to be noticed', based on one of Keller's (1994) 'maxims of action'. Extravagance has recently gained increasing attention in the literature, especially in studies on language change within the framework of Diachronic Construction Grammar (e.g. Petré 2017, De Wit et al. 2020, Kempf & Hartmann 2022, Baumann & Mühlenbernd 2022). As argued by Ungerer & Hartmann (2020), extravagant expressions are typically characterised by a combination of features, including the use of imaginative and vivid language, but also deviations from linguistic norms and expectations. Both of these characteristics – imaginative language and norm violations – feature frequently in snowclones. This is illustrated by the examples in Table 2, which display extravagant formal and/or functional features at different levels of linguistic analysis, ranging from clausal syntax to prosodic structure to individual lexical items.

We would thus argue that, when using snowclones, language users are guided both by the maxim of extravagance ('talk in such a way that you are noticed') and by the

Snowclone	Example	Comment
Striking formal features		
[*Have* X, *will travel*]	*Have tux, will travel*	Parallelism and ellipsis
[*All your* X *are belong to us*]	*All your history are belong to us*	Syntactic violation
[X *2: Electric Boogaloo*]	*Breakin' 2: Electric Boogaloo*	Prosody (rhythm and rhyme)
[*I* ♥ X]	*I* ♥ *NY*	Multimodality
Striking functional features		
[X *BE the new* Y]	*Thursday is the new Friday*	Incompatibility/paradox
[*the mother of all* X]	*the mother of all battles*	Hyperbole
[*Save an* X, *ride a* Y]	*Save a horse, ride a cowboy*	Lexical ambiguity
[*I, for one, welcome our* X *overlords*]	*I, for one, welcome our new insect overlords*	Register/style (pragmatic indexing)

Table 2
Examples of extravagant formal and functional features in snowclones.

604

seemingly opposing maxim that Haspelmath (1999: 1055) postulated on the basis of Keller (1994), the maxim of conformity ('talk like the others talk'). On the one hand, the creative reuse of stylistically striking patterns, in combination with ever changing slot fillers, allows language users to portray themselves as competent and innovative users of their language. At the same time, language users limit this creative variation to the open slots of the constructions, while simultaneously using their fixed formulaic elements to allude to a shared stock of (pop-)cultural knowledge. This may be even more so the case for snowclones that are prevalent among certain cultural groups but opaque to other language users; not everyone may be familiar with the cultural references that underlie examples like [X *2: Electric Boogaloo*] in Table 2. Snowclone users can thus signal their status as members of an in-group by demonstrating that they are familiar with the cultural 'common ground' of the group.

In summary, by combining previous suggestions from the literature with some novel extensions, we propose a definition of snowclones that rests on three key criteria:

Snowclones are a class of partially filled constructions characterised by

  (i) the existence of an (alleged) lexically fixed source construction that is culturally salient and has sufficiently high token frequency to serve as a template for snowcloning;
 (ii) productivity (operationalised via type frequency), i.e. the extension of the pattern to new instances via lexical substitutions in one or more variable slots;
(iii) distinctive ('extravagant') formal and/or functional characteristics that function as markers of linguistic innovation and increase the pattern's memorability.

Importantly, we conceive of these criteria as prototypical characteristics of the category 'snowclones', rather than as strictly necessary and sufficient conditions. This means that while we do expect snowclones to display all three characteristics to a certain extent, the relative weights of the criteria may differ among individual examples. For instance, as we will illustrate in Section 4.3, [X *BE the new* Y] is attested with an extensive set of slot fillers, thus satisfying criterion (ii) of our definition particularly clearly. As mentioned above, however, its lexical source – criterion (i) – is less clearly established than in the case of other snowclones, with speakers being potentially less aware of its putative origin in fashion slogans like *pink is the new black*. Similarly, we will suggest in the following sections that [X *BE the new* Y] can be regarded as semantically extravagant – criterion (iii) – due to the fact that it typically juxtaposes two semantically incompatible or at least nontrivially related concepts (e.g. *antifascism is the new fascism*). Nevertheless, the degree of extravagance is arguably lower than in the case of stylistically marked snowclones like [*I, for one, welcome our* X *overlords*]. This example also points to a potential tension among the three criteria above. As well-established snowclones like [X *BE the new* Y] increase in productivity, they may partially emancipate themselves from their source construction and simultaneously lose some of their perceived extravagance. Alternatively, they may have been perceived as less extravagant from the start, which then facilitated their spread (see also Section 6 for outstanding research questions about the 'life course' of snowclones).

605

In the case studies in Section 4, we will illustrate the three characteristics discussed above by conducting detailed corpus-based analyses of two selected snowclones. Our focus will be particularly on the degree of productivity and semantic variability of the respective patterns (criterion (ii) above), but we will also look for evidence of their lexical source (criterion (i)) as well as their extravagant characteristics (criterion (iii)).

4. CASE STUDIES

In this section, we present case studies of two English snowclones: [*the mother of all* X] and [X *BE the new* Y]. We chose these constructions for two reasons. First, they have been mentioned as typical examples of snowclones in the previous literature (Zimmer 2007, Traugott & Trousdale 2013: 150). Second, as suggested by, for instance, Zimmer (2006), they are particularly frequent members of their category. This impression is confirmed by our corpus analysis (see below), which is based on several thousand occurrences of each pattern in a large web corpus.[3]

The goal of our analysis is to examine to what extent our two selected snowclones display the three characteristics outlined in Section 3. To do so, we analyse the range and semantic type of the lexical items that are attested in the open slots of the snowclones. With respect to the first criterion of our definition, we will assess whether the slot fillers that combine most typically with the snowclones provide evidence of a (putative) lexically fixed source for each pattern. Regarding the second criterion, we will explore the degree of productivity of our two patterns, aiming to characterise both their semantic range as well as the extent to which they are semantically constrained. We will conduct the analysis both on a global level (e.g. using type/token ratios) and on the level of specific semantic clusters that comprise similar slot fillers. In the case of [X *BE the new* Y], we will additionally assess the relationship between the two open slots (X and Y). This can not only provide us with further insights about the semantic profile of the snowclone, but it may also reveal whether the snowclone has spawned productive sub-constructions that only contain a single open slot. Finally, we will look, at least tentatively, for signs that the two snowclones encode features of extravagant language, thus addressing the third criterion from Section 3. While there are no agreed-upon ways of 'measuring' extravagance (see Ungerer Hartmann 2020), the semantics of the slot fillers may provide at least implicit evidence of whether the snowclones feature inconspicuous semantic profiles, or whether they display striking and imaginative characteristics.

---

[3] We also ran preliminary analyses to compare the frequencies of [*the mother of all* X] and [X *BE the new* Y] to other snowclones. We searched the ENCOW corpus (see Section 4.1) for all snowclones included in Tables 1 and 2 (see Section 3). As expected, none of them are attested as often as the two snowclones investigated here. While some of the patterns (e.g. [X *are from Mars,* Y *are from Venus*]) are still in the range of 1,000-2,000 attestations, others occur only a handful of times in the corpus and could not be subjected to quantitative analysis.

606

4.1 *Data and methods*

For both snowclones, we queried three different corpora: The Corpus of Historical American English (COHA, Davies 2010), the Corpus of Contemporary American English (COCA, Davies 2008), and the web corpus ENCOW16A (Schäfer 2015). While the data from the two former as well as the ENCOW data for [*the mother of all* X] were taken into account exhaustively, we worked with a sample of 5,000 instances for the ENCOW data for [X *BE the new* Y]. As we added semantic annotations to the data, this was more feasible than using the full set of 19,818 hits. As the ENCOW corpus is distributed in the form of randomized sentence shuffles (Schäfer & Bildhauer 2012), we simply used the first 5,000 attestations.

False hits were manually excluded (see the individual case studies for the criteria according to which we identified the instances of each construction). Table 3 shows the number of hits (excluding false positives) for each construction in each corpus, along with information about corpus size and the time periods covered by the corpora.[4] As the data from COHA are quite sparse (which can be seen as an indicator that both patterns are relatively recent innovations), we will mainly focus on the COCA and ENCOW data and only use COHA for a qualitative assessment of early examples.

Before we turn to the case studies, we briefly outline the methodology employed for both studies. For each pattern, we will first look at the 'etymology' of the snowclone, taking into account data from COHA and COCA. A quantitative analysis of the patterns' frequency development is combined with a qualitative look at the early data with regard to the semantics of the slot fillers. We then rely on the larger datasets from ENCOW to conduct detailed quantitative analyses of the constructions' synchronic use. For a deeper understanding of the semantic tendencies that the patterns exhibit, we investigate their open slots using simple frequency counts as well as two explorative methods, both of which have come to be widely

Corpus	No. of Tokens	Time Period	Hits for [*the mother of all* X]	Hits for [X *BE the new* Y]
COHA	400m	1810-2009	88	6
COCA	560m	1990-2017	211	82
ENCOW	16.8bn	Web corpus	4,127	3,848*

Table 3
Overview of the corpora and the number of hits. *The value for [X *BE the new* Y] in ENCOW is the number of true positives out of a sample of 5,000 instances.

[4] The version of ENCOW used for the present study was compiled in 2014, which is thus the *terminus ante quem* for all data it contains. Given the nature of the corpus, it is often not possible to pinpoint the publication date of a text that was crawled for the corpus, let alone the time when it was actually written. All in all, however, it seems reasonable to assume that most texts in the corpus are from the early twenty-first century (which matches our subjective impression as frequent users of the corpus).

607

used in constructionist approaches to language variation: collostructional analysis on the one hand and distributional semantics on the other.

Collostructional analysis is a family of methods that allows for quantifying the relationship between constructions (see Stefanowitsch 2013 for an overview). We will employ two of these methods here, simple and covarying collexeme analysis (with the latter only being used in the second case study):

- Simple collexeme analysis (Stefanowitsch & Gries 2003) identifies association patterns typically between lexical items and syntactic constructions (but see Hartmann 2014 and Smirnova 2021 for morphological application variants). For each lexical item i in a syntactic construction C with one open slot, a cross-tabulation test is computed over a 2×2 table containing (a) the frequency of i in C, (b) the frequency in which C is attested with any other lexical item $\neg i$, (c) the frequency of the lexical item i in all other constructions $\neg C$, and (d) the frequency of all other lexical items $\neg i$ in all other constructions $\neg C$. In doing so, one can identify lexemes that occur with above-chance frequency in a construction. These are called 'attracted collexemes', while lexemes that occur much less often than would be expected at chance level are called 'repelled collexemes'. For example, Stefanowitsch & Gries (2003) show with the help of a simple collexeme analysis that for the [X *waiting to happen*] construction, words with negative prosody like *accident* or *disaster* are identified as strongly attracted collexemes.
- Covarying collexeme analysis (Stefanowitsch & Gries 2005) investigates association patterns between two open slots in the same construction. Given a construction with two open slots s_1 and s_2, for each combination of lexical items l_1 and l_2, a cross-tabulation test is computed over a 2×2 table containing (a) the frequency with which l_1 occurs in the first open slot s_1 and l_2 in the second open slot s_2, (b) the frequency with which any other lexeme $\neg l_1$ occurs in the first open slot s_1 while the second open slot s_2 is filled with l_2, (c) the frequency with which l_1 occurs in s_1 while s_2 is filled with any other lexeme $\neg l_2$, and (d) the frequency in which both s_1 and s_2 are filled with other lexemes $\neg l_1 / \neg l_2$. Using covarying collexeme analysis, Stefanowitsch & Gries (2005) examine the two verbal slots of the so-called *into*-causative [V *someone into* V*ing something*], showing that *fool* and *think* are the covarying collexemes that are most strongly attracted to the construction (e.g. *We must not fool ourselves into thinking that…*).

All collostructional analyses reported in this paper were performed using Flach's (2021) package *collostructions* for R (R Core Team 2020). We use the log-likelihood ratio G^2 (Dunning 1993) as our main association measure. There are, however, two caveats. First, G^2 is a bidirectional association measure, i.e. it quantifies the mutual attraction between the construction and its collexemes but does not distinguish the direction of the association (Gries 2019: 387). For this reason, we additionally calculate delta P (construction-to-word), a unidirectional measure of attraction from

the construction to its collexemes. Second, like other measures such as the Fisher-Yates Exact Test used in the early applications of collostructional analysis, G^2 conflates pure association with sample size, i.e. the frequency of the collexeme. As Gries (2019: 389) points out, this can actually be quite useful, especially if the predominant goal of the analysis is to obtain a one-dimensional ranking of items in a construction. However, Gries concedes that it also makes sense to disentangle both dimensions. For this reason, we also compute the log odds ratio as an additional measure of 'pure' association independent of frequency. In doing so, we follow the emerging consensus among users of collocation-based methods to ideally draw on a combination of different measures, as each of them captures different relevant dimensions (see e.g. Schneider 2020).

Apart from collostructional analysis, we use semantic vector spaces. This method has become more and more popular in recent years under the heading of 'distributional semantics' (see e.g. Levshina & Heylen 2014, Levshina 2015, Hilpert & Perek 2015, Perek 2016, Perek & Hilpert 2017, De Pascale 2019, Hilpert & Perek 2022). It allows for quantifying the similarities and dissimilarities between words based on their collocates. For example, a word like *summer* may frequently co-occur with *spring* or *height*, while a word like *disaster* may occur with *victim* or *environmental*.

We use an implementation of word2vec (Mikolov et al. 2013), which has become 'one of the most famous and successful word-embedding schemes' (Chollet & Allaire 2018: 177). The term *word embeddings* simply refers to word vectors, as the word vectors used in this approach are created by embedding the vocabulary in a real vector space R^d (d, indicating the number of dimensions, see Levy 2022: 334). More specifically, we use Schmidt & Li's (2022) R package *wordVectors* to train and explore our word2vec models. The model was trained with the first of the 17 downloadable sets of ENCOW sentence shuffles[5], comprising around 600 million sentences, using a five-word window and a skip-gram approach with negative sampling.[6] To visually represent the results, we use t-distributed Stochastic Neighbourhood Embedding (t-SNE, van der Maaten & Hinton 2008). Like other well-known methods such as Multi-Dimensional Scaling (MDS, Wheeler 2005), t-SNE is a technique for dimensionality reduction, i.e. it transforms configurations of at least $n + 1$ items into fewer than n dimensions. In our case, this results in a representation of semantic similarities in a two-dimensional space. t-SNE has been argued to be particularly well-suited for the visualization of high-dimensional datasets (Desagulier 2017: 254) and to reveal more clear-cut clusters than MDS (see e.g. De Pascale 2019: 202). For implementing t-SNE, we use the R package *Rtsne* (Krijthe 2015), which uses the Barnes-Hut implementation of t-SNE introduced in van der Maatenn (2014).

---

[5] See https://web.archive.org/web/20210924132944/https://corporafromtheweb.org/encow16/ for more information on the downloadable samples (the 'AX' dataset in the right-hand column). The samples themselves are available after registration on https://www.webcorpora.org/. Both links last checked March 2, 2023.

[6] See the supplementary material for further technical details. For more details on the inner workings of word2vec, see e.g. Jurafsky & Martin (2023: Chapter 6.8).

609

Regardless of the concrete method, dimensionality reduction allows for, metaphorically speaking, charting the semantic territory of each construction. As the number of hits (and the overall corpus size) is much higher for ENCOW than for the other two corpora, our semantic vector-space analysis is based exclusively on ENCOW.

4.2 *[the mother of all X]*

Our first case study is devoted to the snowclone [*the mother of all* X], which is used to signal prototypicality or abundance. According to dictionaries of contemporary English, it denotes 'an extreme example of something'[7] or 'something regarded as the biggest, most impressive, or most important of (its kind)'.[8] The lexical source of the snowclone is usually attributed to former Iraqi dictator Saddam Hussein, who used the term 'mother of all battles' in a televised speech on the eve of the First Gulf War (Ferguson 2019: 202). Previous metaphorical uses indicate linear descent. For instance, Lass (2015: 46) cites Isidor of Seville's assertion that Hebrew should be considered the 'mother of all languages and literatures'. However, the use of [*the mother of all* X] in a genealogical sense, even if used metaphorically, cannot be considered a snowclone in our view as these instances constitute relatively straightforward cases of personification. This is why instances like the ones in (2) were discarded. In addition, false hits that obviously do not correspond to the syntactic target structure were filtered out manually, e.g. *The judge acquitted Merlin's mother of all charges* (ENCOW).

(2)   (a)   'X is a literal or metaphorical descendant of its antecedent': *stem cells (which are the mother of all cells – 'mother cells'!)* (Cosentino Fabius: First Word, COCA)

   (b)   'X is brought about by its antecedent': *Necessity is the mother of all innovation* (04884707b18656d2a12f51ce7d55e29bf488, ENCOW)

   (c)   'X is supported/protected by its antecedent': *The banks had failed, but the government had stepped in. It became the mother of all banks* (7ba947acf40e70ed0d1c79f2aeff99b0505b, ENCOW)

Starting with our corpus results, we will first examine some simple measures that are based on the type and token counts for [*the mother of all* X]. These measures provide an overall impression of the snowclone's productivity. Table 4 summarises the number of tokens, types and hapax legomena attested in COCA and ENCOW. Moreover, Figure 1 plots the temporal development of (a) the token and type frequency, (b) the type-token ratio, and (c) the so-called potential productivity (Baayen 2009), i.e. the proportion of hapax legomena to the total number of tokens belonging to the pattern, in COCA between 1990 and 2017.[9] Together, the data

---

[7] https://dictionary.cambridge.org/dictionary/english/mother-of-all-sth (last checked March 2, 2023)

[8] https://www.collinsdictionary.com/dictionary/english/the-mother-of-all_1 (last checked March 2, 2023)

[9] The token frequency indicates the frequency for each year represented in COCA. It was calculated by, for each year, dividing the total number of attestations of [*the mother of all* X] by the total

610

	Tokens	Types	Hapax Legomena
COCA	211	160	138
ENCOW	4127	1669	1092

Table 4
Number of tokens, types, and hapax legomena attested for [*the mother of all* X] in COCA and ENCOW.
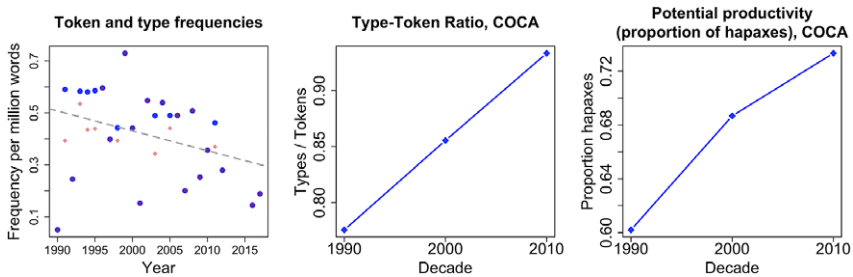

Figure 1
Changes in (normalised) token frequency (big blue dots) and type frequency (small red squares), type-token ratio, and proportion of hapax legomena for [*the mother of all* X] in COCA.

illustrate that the snowclone has been quite productive since its beginnings. Both the type-token ratio and the proportion of hapax legomena have been consistently high; they even seem to have increased somewhat over the period of investigation. This increase in potential productivity should not be overestimated as the measure of potential productivity partly depends on token frequency (see e.g. Gaeta & Ricca 2006, Hartmann 2018). Nevertheless, while the sparse diachronic data should be interpreted with some caution, they suggest that the snowclone has overall maintained high levels of productivity.

As a second step, we use a simple collexeme analysis to identify the lexemes that are strongly attracted to the open slot of [*the mother of all* X]. This way, we can gain a more detailed understanding of the semantic range covered by the snowclone. Tables 5 and 6 report the top 10 simple collexemes of the construction in COCA and ENCOW, respectively.[10] The collexemes are ranked by our main association measure, the log-likelihood ratio G^2. As explained in Section 4.1, we also report two additional measures, the unidirectional delta P (construction-to-word) and the

---

number of running tokens in the corpus. With around 20 million words per year, the COCA corpus is relatively well-balanced. For the type-token ratio and the potential productivity, we chose a more coarse-grained periodisation and worked with decades instead of years to show the overarching trend. The type-token frequency was calculated by dividing the number of [*the mother of all* X] slot filler types by the number of slot filler tokens; the potential productivity was calculated by dividing the hapax legomena, i.e. slot fillers attested only once, by the total number of instances of [*the mother of all* X].

[10] See the supplementary material for interactive tables containing the full lists of collexemes.

611

Collexeme	Frequency in Cxn	Total Frequency	G^2	ΔP Cxn-to-Word	Log Odds Ratio
battle	21	45078	187.10654	0.09908	2.39941
storm	5	31113	33.59615	0.02339	1.93326
jam	3	9363	24.23168	0.01412	2.25445
stickup	1	8	20.11794	0.00474	4.97861
op	2	2645	19.5654	0.00945	2.6554
bomb	3	26793	17.99518	0.01395	1.7977
squeezer	1	34	17.12335	0.00474	4.32863
headache	2	6169	16.19178	0.00942	2.28745
comb-over	1	60	15.97455	0.00474	4.07915
briefer	1	87	15.22631	0.00474	3.91665

Table 5
Results of the simple collexeme analysis for [*the mother of all* X] in COCA, sorted by G^2.

Collexeme	Frequency in Cxn	Total Frequency	G^2	ΔP Cxn-to-Word	Log Odds Ratio
battle	148	577746	1113.36172	0.0361	2.07358
hangover	62	18482	783.25739	0.01525	3.18476
bubble	50	142649	406.05799	0.01222	2.20203
crisis	60	556466	347.64395	0.01446	1.6902
bailout	28	46997	256.80882	0.00686	2.43353
headache	29	95519	227.08792	0.00708	2.14046
adventure	34	274370	206.06747	0.00821	1.75063
cock-up	15	2875	202.59408	0.00369	3.38301
storm	28	246347	164.91399	0.00675	1.7138
bomb	27	239152	158.66175	0.00651	1.71106

Table 6
Results of the simple collexeme analysis for [*the mother of all* X] in ENCOW, sorted by G^2.

frequency-independent log odds ratio. As can be seen from the results, the delta P values yield a similar ranking as the one obtained by G^2. For the log odds ratio, there are some more differences in the ranking of individual collexemes. However, since the values are nevertheless consistently high, and since our interpretation relies on an overall assessment of the typical semantics of the collexemes, rather than on their exact ranking, we restrict our further discussion to the G^2 values.

Two key observations can be made based on these data. First, in both corpora, the most strongly attracted collexeme is *battle*, thus hinting at the lexically fixed source of the snowclone (see Saddam Hussein's quotation above). While the pattern has been extended to various other slot fillers, none of these have reached the same frequency as the source construction. This suggests that the snowclone has not fully emancipated from its lexical source, which continues to be culturally salient for speakers.

612

Figure 2
Semantic vector-space analysis of the lemmas occurring in [*the mother of all* X] in ENCOW.

Second, the top 10 collexemes indicate certain semantic preferences among the slot fillers. In particular, [*the mother of all* X] seems to combine primarily with abstract concepts that have a negative semantic prosody (e.g. *storm, crisis, headache, hangover, bailout*). Moreover, the slot fillers include a number of everyday concepts (e.g. [*traffic*] *jam, headache*) and also colloquial terms (e.g. *cock-up*). The fact that several of the top collexemes describe dangerous or otherwise unpleasant events indicates that speakers use the snowclone in situations in which they are emotionally strongly involved (which is often seen as one of the defining criteria of extravagance; see Section 5 for discussion).

To further probe the semantic range of the slot fillers, we conduct a word2vec semantic vector-space analysis based on the ENCOW data (as described in Section 4.1). The results are shown in Figure 2. As outlined above, t-SNE allows for representing distances in two-dimensional space. Hence, items that are displayed close to one another are identified as similar, based on their collocates, while items that are farther away from each other are rather dissimilar. The font size of the lexemes represents the frequency with which they are attested in the [*the mother of*

all X] construction.[11] To avoid too many overlapping words, the package *ggrepel* (Slowikowski 2021) was used, which repels individual words from overlapping items – hence the lines in the plot, which indicate the actual position of the repelled words.

A first impression of Figure 2 is that the collexemes are quite broadly dispersed over the semantic space, again testifying to the high degree of productivity and semantic flexibility of the snowclone. Nevertheless, on a closer look, we can distinguish several groups of semantically similar items in the diagram. First, *battle* stands out as the most frequent item in a group of lexemes in the middle and to the right of the plot, many of which relate to the military domain (*bomb, storm, war*). A second group in the (lower) left of the plot consists of abstract terms referring to organisations or institutions (*state, union*), person collectives (*group, panel*), and events (*campaign, revolution*). The third major group clusters in the upper left of the plot, with *headache* as its most frequent member. It contains words with a rather negative semantic prosody, many of which belong to the domain of health conditions or risky health behaviour (*hangover*, *overdose*, *abscess*). Finally, at the top of the plot, we find a group of terms related to food, such as *pork, pizza* and *sauce*, but also *dinner* and *restaurant*.[12]

Finally, we use our semantic vectors to test how closely the meanings of the collexemes are related to the meaning of the fixed element *mother*.[13] This is relevant insofar as the conflict between the prototypical meaning of *mother* and its specific (non-literal) use in the snowclone arguably contributes to the extravagance of the construction. We calculate the semantic distance between each slot filler and *mother* by subtracting their cosine similarity value from 1. As the cosine similarity ranges from -1 to 1, the cosine distance ranges from 0 (completely identical) to 2 (completely opposite). As the histogram in Figure 3 shows, few slot fillers are closely related to the core meaning of *mother*. Rather, they are typically fairly distinct. In *the mother of all pizzas* and *the mother of all asteroids*, for example, the semantic distance between *mother* and the X element is quite large (0.83 and 0.84, respectively). This suggests that, when using the snowclone, speakers shift *mother* away from its original lexical sense and use it to create a new idiomatic meaning.[14]

---

[11] To make the individual items more readable, we have slightly increased the font size for the print version. See the online supplementary material for a version of the plot that shows the frequency differences more clearly.

[12] The example of *jam*, which appears in close proximity to food items here, illustrates an inherent limitation of our approach. In our data, *jam* actually only occurs within the compound *traffic jam*. Both in the case of [*the mother of all* X] and [X *BE the new* Y], however, most of the slot fillers are one-word items, which is why we semi-automatically extracted the head words of compounds and phrases that were used as slot fillers. We would still argue that the semantic vector-space analysis yields valuable results as long as they are interpreted with caution.

[13] Thanks to an anonymous reviewer for suggesting this.

[14] To better evaluate the size of the semantic distances in Figure 3, we also compared them with the distances between *mother* and a random selection of 10,000 nouns from the corpus. As shown in the supplementary material, the distributions of cosine distances peak around the same value. This highlights that the snowclone combines flexibly with a range of slot fillers that are not
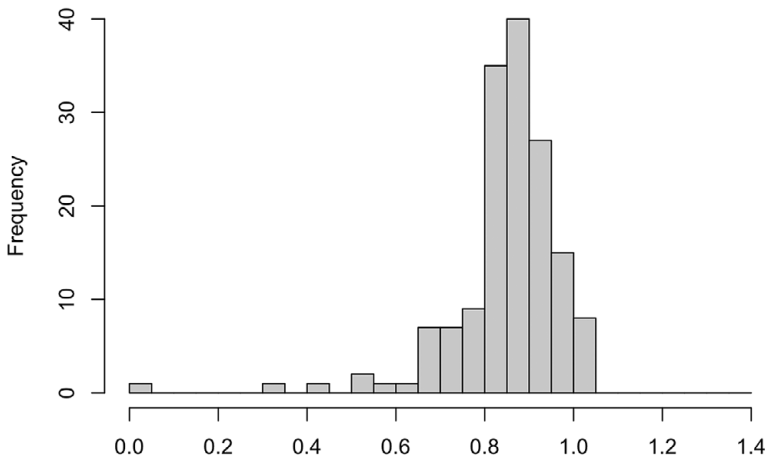
614

Cosine distance between *mother* and all X items

Figure 3
Cosine distance between *mother* and the items in the X slot of [*mother of all* X] as per the word2vec distributional-semantic analysis.

Together, these findings suggest that [*the mother of all* X] combines with a broad range of slot fillers. Nevertheless, this diversity does not seem unlimited, as many of the slot fillers still cluster around certain semantic domains. In addition, our collostructional analysis indicates that the most strongly attracted collexemes have a negative semantic prosody, suggesting that the pattern is still somewhat tied to the semantics (and pragmatics) of its original lexical source (*the mother of all battles*). We will further discuss the interplay between productivity and semantic constraints in Section 5.

4.3 *[X BE the new Y]*

Our second case study is concerned with the snowclone [X *BE the new* Y], which is 'used to state X is now fashionable (or common), where Y was before'.[15] For example, *smart is the new young* (CBS_SunMorn, COCA) expresses that intelligence has come to play a greater role in a context where young age previously used to be a more decisive factor. The origins of the snowclone are usually attributed to fashion slogans from the 1970s and 1980s (e.g. *pink is the new black*), even though it is not clear when exactly and with what colour terms the snowclone was first attested (see Zimmer 2006). As mentioned in Section 3, the

---

closely related to the core meaning of *mother*. Moreover, the comparison with the random sample shows that cosine distances larger than 1.2 are exceedingly rare in our data, suggesting that the values in Figure 3 lie towards the upper end of the spectrum.

[15] https://www.wiktionary.org/wiki/Appendix:Snowclones/X_is_the_new_Y.

615

lexical source of the construction is therefore less clearly established than for other snowclones.

In order to identify instances of the pattern, we searched the corpora exhaustively for the string *is/are the new* preceded and followed by a token tagged as either an adjective or a noun.[16] Attestations in which the pattern is used to refer to an identity or a class inclusion relation (Glucksberg & Keysar 1993: 412), as illustrated in (3), were manually discarded; this also includes metaphorical instances such as (3-c). Unclear cases were marked as such but not taken into account in the final analysis. For example, nearly all instances of [X *is the new religion*] can be seen as doubtful cases as it is unclear whether they are used in the sense of 'X is (like) a new type of religion', thus constituting a (metaphorical) class inclusion relation, or rather in the sense of 'X has taken the place of religion' (the snowclone meaning).

(3) (a) *Penny Adkins is the new manager of the Imaging unit of Central Services.* (Denver Post, People on the Move, 1997, COCA).
  (b) *Daniel Craig is the new James Bond* (0fb25d98cec169aa250051 d8144bb34c11ab, ENCOW)
  (c) *after the French macarons, cupcakes are the new scene stealers this year* (251fe7457256dc055976e77fb3fcddefb593, ENCOW)

Before delving into the details of the corpus analysis, a few qualitative observations can be made that already illustrate some of the productivity and extravagance of [X *BE the new* Y]. At its core, the snowclone describes a comparison between the two concepts in the X and the Y slot. Given that, as Cummins (2019: 144) points out, 'pretty much everything is like pretty much anything else in at least some identifiable respect', the snowclone thus invites speakers to draw original and innovative comparisons. In this sense, the snowclone bears similarity to other constructions like [X *BE the* PROPER NAME *of* NP], as in *She is the Einstein of Cognitive Linguistics*. These constructions have been described as metaphorical (see e.g. Glucksberg & Keysar 1993, Grady 1999: 98) and can also be seen as conceptual blends in the sense of Fauconnier & Turner (2002). In these cases, 'the metaphor provides a high degree of implicative elaboration' (Glucksberg & Keysar 1993: 421). Glucksberg & Keysar's example *Xiao-Dong is a Bela Lugosi*, for instance, builds on the hearer's knowledge that Bela Lugosi was an actor best known for his portrayal of Dracula. In using him as a reference point, he is basically 'typified', and depending on the context (and their cultural knowledge), the hearer

---

[16] It has been claimed that the adjectival pattern [ADJ$_1$ *BE the new* ADJ$_2$] is the construction that gave rise to [X *BE the new* Y] and that the latter was only later extended to nouns (see e.g. Traugott & Trousdale 2014: 272; also see the blog posts by Zimmer 2006, Zwicky 2006a). However, the [N$_1$ *BE the new* N$_2$] pattern has been attested in COHA already in the late nineteenth/early twentieth century, as the following examples show: *'To-day', he sings, 'Milan is the new Athens!'* (Julia Cartwright Ady, Beatrice d'Este, Duchess of Milan, 1475-1497, 1899, COHA); *The education department of the Government of India is the new Frankenstein, and the higher education is its monster* (The Atlantic Monthly, October 1913, COHA).

616

will identify the salient features that the speaker refers to when comparing Xiao-Dong to Bela Lugosi.

Similarly, instances of [X *BE the new* Y] such as *Spielberg is the new Capra* (Love Walked In, COCA) require a 'high degree of implicative elaboration' in that the hearer has to establish the *tertium comparationis* that connects the two directors (which can of course be more than a single feature). Other cases, such as *Android is the new Windows, and Apple is about to see the 1990s all over again* (b6137beed83ff4cc69f6bf4a58fcfdd9a9b4, ENCOW), tap into an even more complex configuration of shared encyclopaedic knowledge. In this example, hearers need to relate the competition between Apple and Microsoft and the success of Microsoft Windows in the 1990s to the present-day competition on the smartphone market. Quite frequently, the comparisons are non-obvious enough that speakers decide to state the *tertium comparationis* explicitly, as shown in (4).

(4) (a) America is the new Rome. We like physical sports. (75daf596dc2196 a8348667153ba4e8e74f27, ENCOW)

(b) If depression was the disease of the moment in the 1990s, anxiety is the new depression. (5a2d377a41ce4ff55fdf3988091c3e411163, ENCOW)

(c) Google glasses are the new segway – useful if your company pays you to use it but no consumer who is not a creeper is going to buy one. (7f51314f251f58348c69ac6d5ffcbff182cb. ENCOW)

Together, these examples suggest that [X *BE the new* Y] is a productive pattern that tends to attract extravagant combinations of slot fillers. In the following, we will examine to what extent these initial impressions are confirmed and extended by our quantitative corpus analysis.

As in our first case study, we start by examining the token and type frequencies of the snowclone. However, given the low number of hits in COCA, we do not assess the diachronic development of their distribution. Table 7 summarises the number of tokens, types, and hapax legomena in both corpora. The latter two are calculated for pairs of X and Y elements as well as separately for the X slot and Y slot. The high type frequency relative to the number of tokens (which corresponds to type-token ratios of 0.87 and 0.73 for the X-Y pairs in the two corpora) suggests that [X *BE the*

	COCA			ENCOW		
Tokens	82			3848		
	X-Y pairs	X	Y	X-Y pairs	X	Y
Types	71	67	52	2805	2000	1651
Hapax legomena	64	60	47	2479	1505	1241

Table 7
Number of tokens, types, and hapax legomena attested for [X *BE the new* Y] in COCA and ENCOW.

617

Collexeme	Frequency in Cxn	Total Frequency	G^2	ΔP Cxn-to-Word	Log Odds Ratio
green	91	1227752	500.12	0.0249	1.62506
pink	40	197699	298.4	0.01112	2.05815
black	67	1743023	282.79	0.01791	1.33763
small	77	3367848	250.03	0.01989	1.11246
white	55	1884617	203.29	0.01447	1.21718
grey	27	214880	175.89	0.00746	1.8522
blog	38	971381	161.37	0.01017	1.34429
brown	32	700039	145.39	0.00862	1.4123
facebook	27	435778	138.5	0.00735	1.54506
data	51	3013370	137.4	0.01277	0.98011

Table 8
Results of the simple collexeme analysis for the X slot of [X *is the new* Y] in ENCOW, sorted by G^2.

new Y] is a highly productive pattern. This is also supported by the number of hapax legomena among X-Y pairs, which make up between 64 and 78% of all attested tokens. Interestingly, a comparison between the type frequencies of the two open slots indicates that the X slot is overall more productive than the Y slot (type-token ratios of 0.52–0.82 versus 0.43–0.63 across the two corpora). One reason for this may be that one specific Y filler (*black*) is particularly frequent and combines with a range of different X elements (see below).

Following this quantitative assessment of the snowclone's productivity on a global level, we next conduct simple collexeme analyses to gauge the typical semantics of the slot fillers. In this and the following analyses, we only use data from ENCOW because the number of instances in COCA is too small to draw reliable conclusions about the constructional semantics of the pattern. Tables 8 and 9 provide the top 10 most strongly associated collexemes for the X and Y slot, respectively. As in the case of [*the mother of all*], the results are ranked by our primary association measure G^2, but we additionally report delta P values and log odds ratios. Again, while these measures give rise to slightly different rankings, the differences do not affect our overall interpretation, which is why we focus on the G^2-based results in the following.[17]

A first observation based on Tables 8 and 9 is that colour terms are clearly among the most typical fillers in both slots. This provides evidence of the lexical source of the snowclone, which seems to have originated from aesthetic comparisons in the domain of fashion. Interestingly, the X slot displays more variability in the specific colour that is used (e.g. *green, pink, black, grey, white*), and the Y slot is by far most often filled by *black* (and much less frequently by *pink*). *Black* occurs a total of 622 times in the Y slot, thus explaining the above observation that the type frequency among the Y elements is lower than among the X elements. This could

---

[17] As before, the full list of collexemes including interactive tables that allow for ranking the data by delta P or by the log odds ratio are reported in the supplementary material.

618

Collexeme	Frequency in Cxn	Total Frequency	G^2	ΔP Cxn-to-Word	Log Odds Ratio
black	622	1743023	5,503.80	0.18077	2.39894
sexy	32	82298	282.77	0.0093	2.36
oil	57	1095176	278.08	0.01609	1.48654
big	75	2837001	269.42	0.02046	1.19337
nazi	26	116573	200.97	0.00753	2.11931
pink	27	197699	182.49	0.00779	1.90606
jews	29	423586	156.7	0.00825	1.60579
cupcake	14	13211	151.64	0.00408	2.80194
kkk	12	5835	145.88	0.0035	3.09253
objectivity	14	27026	131.64	0.00408	2.49086

Table 9
Results of the simple collexeme analysis for the Y slot of [X *is the new* Y] in ENCOW, sorted by G^2.

indicate that [X *BE the new* Y] originated from a set of related source constructions, which combined *black* with a range of different colours.

Apart from colour terms, the simple collexeme analyses suggest that the construction attracts concepts from the internet and (social) media domain (*blog, facebook, transparency*), as well as group membership terms (*nazi, KKK, jews*). These semantic fields point to specific functional contexts in which the snowclone is used. On the one hand, [X *BE the new* Y] frequently occurs in internet forums and online media. It is therefore not surprising that speakers use the snowclone to draw innovative comparisons between technological developments. Beyond that, the use of group membership terms suggests that the snowclone is also used to characterise (groups of) people in potentially stereotypical and/or pejorative ways.

To test whether these impressions hold across a larger range of slot fillers, and to obtain a more objective measure of semantic similarity, we next analyse the X and Y elements using word2vec. The results are visualised in Figure 4, which again uses t-SNE to group semantically similar slot fillers close to each other in a two-dimensional diagram. The colours indicate the relative frequency of each word in the X or Y slot. Items shown in focal red are attested in the X slot exclusively, items in shown in focal blue are only attested in the Y slot, and items shown in different shades of purple occur in both slots. To keep the plot readable, only items attested at least 10 times in the data are displayed.

Overall, three major clusters can be gauged from the diagram, which correspond quite closely with the three semantic fields we identified in the simple collexeme analyses. The cluster on the right-hand side of the plot mostly contains colour terms (e.g. *black, green, red, pink*). In the lower part, we find internet- and media-related concepts (e.g. *data, blogging, twitter, facebook*), and on the left, we find a cluster comprising social groups, religions, and ideologies (e.g. *republican, islam, racism*).

Combining the results of the simple collexeme analysis and the semantic vector-space analysis, we can observe that [X *BE the new* Y] displays specific semantic preferences, thus covering a set of relatively constrained functional niches. Within

619

Figure 4
Semantic vector space analysis of the slot fillers in [X *BE the new* Y] based on ENCOW.

its limited semantic profile, the construction is highly productive, but it does not seem to display the same degree of semantic flexibility that would be expected from a canonical comparison construction such as [X *is like* Y]. It may be partially due to its constrained applicability, and its association with certain usage contexts (e.-g. internet forums), that the snowclone maintains its innovative and extravagant character.

Having outlined the semantic profile of the individual slot fillers, we now turn to how the X and Y elements relate to one another. As mentioned at the beginning of this section, snowclones like [X *BE the new* Y] provide an interesting test case to study how the semantics of the two open slots interact with each other. This is particularly relevant since it may allow us to characterise, on a quantitative level, what makes the comparisons between the X and Y concepts so innovative, and thus gives the snowclone its potentially extravagant character.

We approach this question using several methodological tools. First, we manu-ally classify all slot fillers in the ENCOW data into coarse semantic categories.
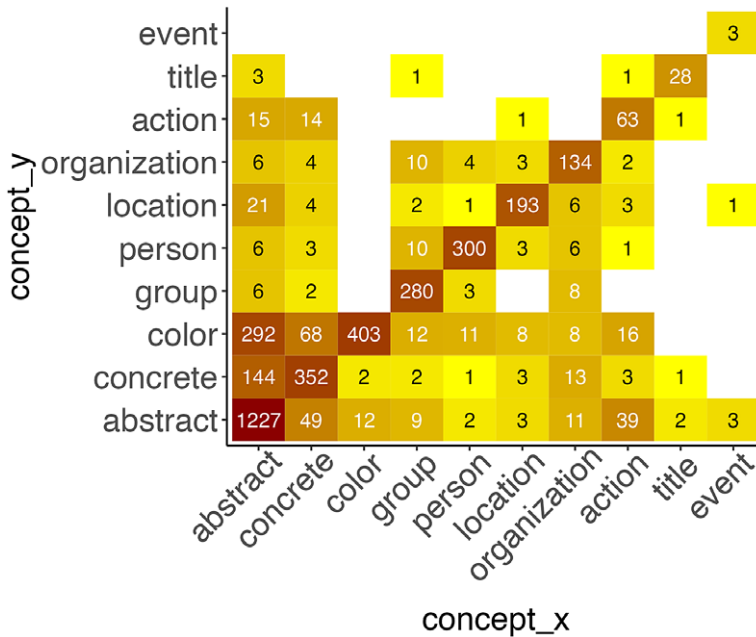
620

Figure 5
Heat map of semantic categories co-occurring in the X and Y slot of [X *BE the new* Y] based on ENCOW
(the darker the colour, the more frequent the combination).

These categories include abstract concepts (e.g. *communism, crazy*), concrete objects (e.g. *banana, SUV*), colours (e.g. *blue, black*), social groups (e.g. *Democrats, Rolling Stones*), persons (*Al Capone, Obama*), locations (*Las Vegas, Mexico*), and organisations (*airline, Google*). We then count how often each semantic category in the X slot co-occurs with each type of Y element. The results are visualised with the help of a heat map in Figure 5.[18]

A first regularity shown by the diagram is that all types of slot fillers occur most frequently with a concept in the other slot that belongs to the same semantic category. This is not particularly surprising. Abstract concepts are typically compared to abstract concepts, locations are compared to locations, and so on. Besides that, however, one other feature deserves attention: abstract X elements are relatively often paired with concrete concepts or colours in the Y slot (see the first column of Figure 5). These combinations occur much more frequently than the opposite pairings of concrete or colour Xs with abstract Ys. This might point to a specific communicative function of [X *BE the new* Y], where it is used to render abstract concepts more tangible by relating them to a more concrete conceptual

---

[18] The colours in the heat map indicate the logarithmised frequency in order to make the visual difference clearer, and the labels show the original frequency values.

621

X Slot	Y Slot	Freq in X Slot	Freq in Y Slot	Combined Frequency	G^2
small	big	77	75	75	720.65
data	oil	51	57	37	283.66
blog	resume	23	28	20	198.98
transparency	objectivity	19	14	14	163.3
muslims	jews	18	29	15	139.99
green	red	91	33	23	138.36
old	new	23	16	13	126.99
strong	skinny	11	13	10	118.28
quiet	loud	8	8	8	114.8
nra	kkk	9	12	9	113.53
thursday	friday	13	20	11	112.15
big	small	8	10	8	104.79
service	marketing	15	26	11	98.04
google	microsoft	27	21	12	96.74
bad	good	7	8	7	96.29
disk	tape	7	8	7	96.29
black	white	67	20	14	92.25
x	y	10	7	7	90.1
atoms	bits	6	6	6	89.55
kindergarten	grade	6	6	6	89.55

Table 10
Results of a covarying collexeme analysis of [X *BE the new* Y] based on ENCOW.

domain. Examples such as *Twitter is the new cigarette* (ca2bc545af6c-be8255bcd36968dc7d675c28, ENCOW) or *The social network is the new production line* (56de85af6ae995d4ecd2f44ea17f77df0ba1, ENCOW) illustrate the impact of social media usage by comparing it to everyday concepts from the domains of recreational consumption and manufacturing.

While the heat map suggests that speakers tend to combine X and Y elements from the same semantic category, the question remains whether the comparisons they draw within each category are innovative (and extravagant) or rather conventional. To investigate this question, we first use another type of collostructional analysis – covarying collexeme analysis – to identify pairs of X and Y elements that most typically occur together in the snowclone. Table 5 summarises the X–Y pairs whose members are most strongly associated with each other. For example, the first row of the table indicates that *small* and *big* are almost always combined when they occur in the snowclone (75 times out of 75 and 77 individual occurrences). This example, along with a few other instances (*old is the new new*, *strong is the new skinny*, and *quiet is the new loud*), illustrates that some of the most typical X–Y pairs in the construction consist of (near-)antonyms. This provides at least tentative evidence that the snowclone is used to draw extravagant comparisons between seemingly incompatible concepts. For illustration, consider paraphrasing the examples with a canonical comparison construction. For example, *small is like*
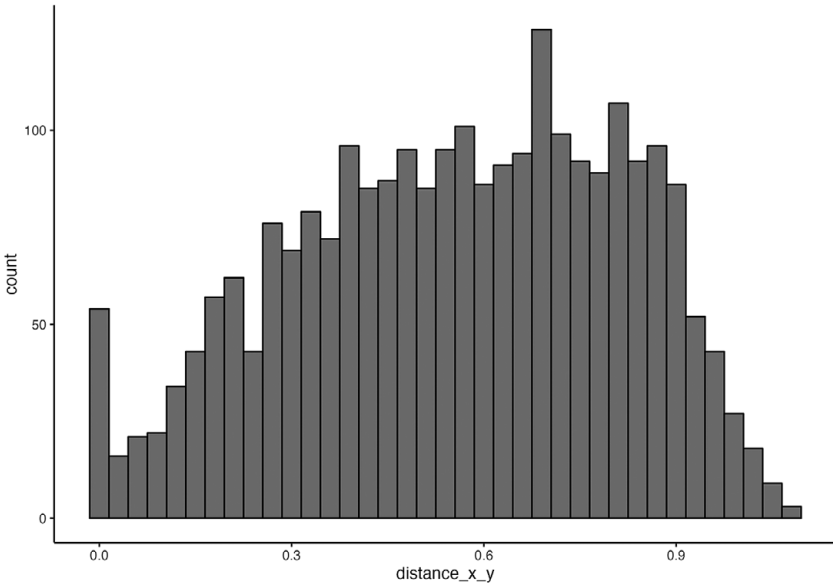
Figure 6
Distribution of semantic distances between X and Y elements in [X *BE the new* Y] based on ENCOW.

big or *strong is like skinny*. Clearly, these sentences sound odd compared with the snowclone examples, suggesting that speakers are more willing to accept unusual combinations when they are presented in the snowclone format.

Finally, we use our semantic vectors to calculate the semantic distance between the X and Y element in each attested pair. This gives us a quantitative measure of how (un)usual the comparisons drawn between the X and Y fillers are. The histogram in Figure 6 summarises how often the different distance values occur in our data (as before, the minimum distance is 0 and the maximum distance is 2). The results suggest that the semantic distances between X and Y occupy a large spectrum, ranging from cases in which X and Y are fully identical, as in (5-a), to items that are semantically quite distinct, such as *Facebook* and *(town) hall* in (5-b).

(5) (a) *The knee-jerk reaction to such a thing would be to call Disney the new Pixar. But the more accurate coronation is that* **Disney is the new Disney**. (ENCOW, c09dbbe27b77df7d2a14ce87819977852806)

(b) **Facebook is the new town hall**, *its censorship is really troubling* (ENCOW, ac437f3e965b48dc8fc66fdea975a7418c32)

Figure 6 also suggests that the semantic distances between the X and Y elements are generally shifted towards the higher end of the spectrum. This claim is, of course, somewhat tentative given that no direct comparison point is available (e.g. a distribution of distance scores for other constructions). Some further examples

623

can, however, help to get a 'feel' for what a semantic distance of around 0.7 (the most frequent value in Figure 6) means. Apart from the examples in (5), other instances that (approximately) illustrate this semantic distance relate X and Y elements such as *truth–speech*, *publishing–literacy*, *innovation–selfishness*, and *history–cooking*. All of these comparisons seem nontrivial and rather creative, thus potentially requiring additional contextual knowledge on the addressee's part. For example, in one attestation, *publishing is the new literacy* is used to compare the growing opportunities to publish one's own work, especially online, with the way in which literacy, once reserved for a small privileged group, spread to the broader public. The fact that such innovative comparisons occur frequently in [X *BE the new* Y] hints at its extravagant character. Rather than comparing just any kind of consecutive trends or fashions, speakers seem to use the snowclone primarily to highlight interesting commonalities between seemingly disparate concepts.[19]

Summing up, our analyses in this section have shown that [X *BE the new* Y] is a highly productive construction, but that many of its uses are still constrained to certain semantic domains. As for the combinations of X and Y elements, users of the snowclone tend to combine polar opposites (*small is the new big*) or cohyponyms (*Thursday is the new Friday*) from the same semantic domain or items from rather different domains (*mean is the new green*), leading to innovative and creative comparisons.

5. DISCUSSION

The results of our two case studies reported in the previous section indicate that both [*the mother of all* X] and [X *BE the new* Y] exhibit all three typical characteristics of snowclones outlined in Section 3. Regarding the first criterion, our corpus results provided clear signs of a lexically fixed source for both patterns – or, in the case of [X *BE the new* Y], arguably a family of source constructions. For [*the mother of all* X], *battle* emerged as the most strongly attracted collexeme of the construction, and [X *BE the new* Y] is associated with colour terms that hint at its origin in fashion slogans of the type *pink is the new black*.

With respect to the second criterion, both patterns display a considerable degree of productivity, while simultaneously not being completely unconstrained in the types of lexemes they combine with. Our results suggest that [*the mother of all* X] typically occurs with words that have a negative semantic prosody, often describing

---

[19] As a reviewer points out, one feature of distributional-semantic methods is that they tend to treat antonyms and cohyponyms like synonyms, as they all usually occur in similar contexts (e.g. Levy 2022: 338; see Nguyen et al. [2016, 2017] for attempts to circumvent this problem). As a result, pairs of antonyms and incompatibles in [X *BE the new* Y] display small semantic distances in our word2vec analysis, even though we have argued above that they are potentially extravagant. This illustrates the strength of combining several methods, as we have done here, each of which may pick up on different types of extravagance. Notably, our vector space analysis indicates that, even without taking antonyms and cohyponyms into account, the semantic distances in [X *BE the new* Y] are quite large, suggesting that the X and Y elements often instantiate different conceptual domains.

military terms or other emergency situations (*war, crisis*) as well as health conditions (*hangover, headache*). Positive uses of the construction are also attested but less common. They are used, for instance, to highlight the quality of food-related items (*restaurant, pizza*). [X *BE the new* Y], however, is somewhat more heterogeneous in that both nouns and adjectives can occur in its open slots. Still, most of its instances centre around certain semantic domains, including colours (*black, green*), group membership and institutional terms (*republican, church*) as well as internet- and media-related concepts (*data, transparency*). The fact that both snowclones display certain semantic preferences suggests that they are still somewhat tied to the semantics (and pragmatics) of their original lexical sources, thus illustrating again the interplay of the first two criteria outlined in Section 3.

Regarding the third criterion, both constructions can be considered extravagant in the sense of Haspelmath (1999), although they seem to draw their extravagance from different sources. As suggested in Section 3, the extravagance of [*the mother of all* X] derives partly from the hyperbolic meaning of the construction: not only does the pattern denote an extreme example of some kind, but it does so in an exaggerated way that often does not correspond to the literal truth (e.g. something that is described as the *the mother of all hangovers* is probably not the most severe hangover to ever be witnessed). This hyberbolic quality aligns well with the evidence from our corpus analysis that the top collexemes of [*the mother of all* X] typically denote emotionally loaded concepts (e.g. *battle, crisis, cock-up*). This suggests that speakers use the snowclone especially in situations of high emotional involvement, in which they want to emphasise their message by extravagant linguistic means. This also links up with previous discussions in the literature, where speaker involvement has been regarded as a defining criterion of extravagance (Petré 2017: 125; see Ungerer & Hartmann 2020 for a critical discussion). Finally, our vector-based results suggest an additional factor that may contribute to the extravagance of [*the mother of all* X]. Most of its slot fillers are semantically quite distant from the typical meaning of *mother*, thus extending the metaphorical use of the concept to new conceptual domains.

[X *BE the new* Y], however, seems to draw its extravagance primarily from the fact that it often involves creative and unconventional comparisons between its X and Y elements. Some of these may be surprising because they are self-explanatory but not necessarily self-evident (*sugar is the new nicotine*), while others need an explicit elaboration of the *tertium comparationis* in order to be understood (*Google glasses are the new segway*). Our analysis of the semantic distances between X and Y fillers provides at least tentative support for the fact that innovative comparisons are the norm, rather than the exception, among instances of the snowclone. Methodologically, these methods provide a potentially novel way of quantifying, in a bottom-up fashion, the extravagant nature of the comparisons encoded by [X *BE the new* Y].

Having discussed to what extent our two constructions exhibit the typical features of snowclones, we now return to another question raised in Section 3, namely how these criteria can be used to set snowclones apart from other partially
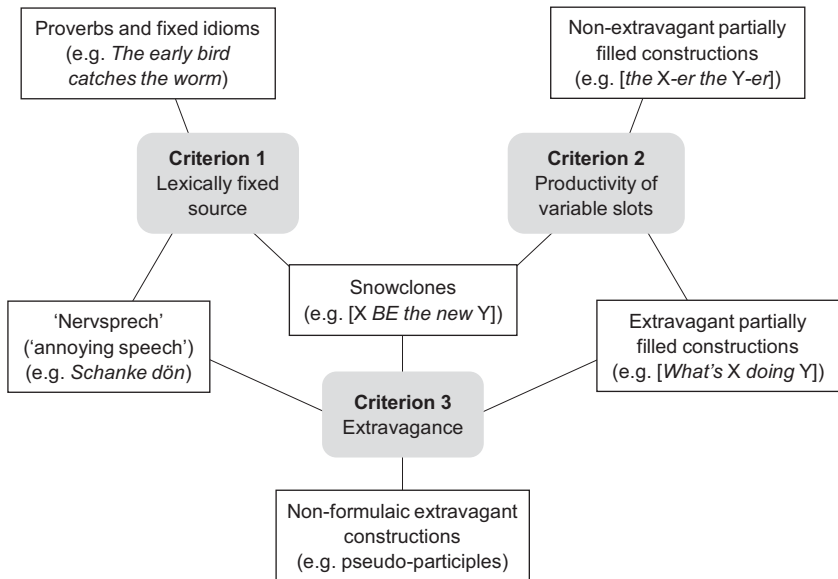
Figure 7
Relationship between snowclones and other types of idiomatic constructions.

similar phenomena. Our view here is that while snowclones share certain features with other types of idiomatic constructions, the *specific combination* of the three criteria we have suggested nevertheless distinguishes them from related phenomena and characterises them as a class of their own. Figure 7 provides a graphic illustration of this idea. Although each typical feature of snowclones is individually also present in other construction types, snowclones are special in that they display all three characteristics at the same time.

Starting with the top left corner of Figure 7, the distinction between snowclones and proverbs (e.g. *The early bird catches the worm*), as well as other lexically fixed or 'substantive' [14] idioms (e.g. *kick the bucket*), is relatively straightforward. While these constructions resemble the lexically fixed source of snowclones, they do not contain open slots that can be filled with variable items, thus failing to fulfil the second criterion of our snowclone definition. It is also doubtful whether all proverbs and fixed idioms necessarily employ extravagant language to convey their content.

Extravagance may, however, be more typical of a related phenomenon for which the German term 'Nervsprech' (roughly: 'annoying speech') has been proposed (Schulze 2013, Finkbeiner 2019). These utterances consist of funny or creative stock phrases which, as a result of their overuse, are thought to have lost their novelty and appeal. Examples include *Tschüssikovsky* (a blend of German *tschüss* 'bye' and *Tchaikovsky*), *Schanke dön* (instead of *Danke schön* 'Thank you'), or *Der Apfel fällt nicht weit vom Pferd* ('The apple doesn't fall far from the horse').

626

'Nervsprech' utterances are usually stylistically marked, often in virtue of their deviation from a more conventional expression. While they may therefore fulfil both the first and the third criterion of our snowclone definition, as shown on the left-hand side of Figure 7, they still lack the open slots that are characteristic of snowclones and are instead relatively invariable.

A parallel argument can be made for the two phenomena shown on the right of Figure 7. Previous work has investigated a variety of other partially filled constructions, such as [*the* X-*er the* Y-*er*] (Culicover & Jackendoff 1999), [*What's* X *doing* Y?] (Kay & Fillmore 1999), and [N *waiting to happen*] (Stefanowitzch & Gries 2003). These constructions have also been discussed under alternative labels, such as 'formal idioms' (as opposed to 'substantive'; Fillmore et al. 1988), 'constructional idioms' (Booij 2002), and 'phraseme constructions' (Dobrovol'skij 2013). All of these partially filled constructions are (partially) productive due to their open slots (see Zeldes 2013 for a detailed discussion), thus fulfilling the second criterion of our snowclone definition. Moreover, while some of the patterns may not be extravagant (e.g. it is doubtful whether speakers would perceive [*the* X-*er the* Y-*er*] as marked, given its degree of conventionality), other constructions do seem to engender special stylistic effects. Compare, for instance, the unusual syntax of [*What's* X *doing* Y?], especially in combination with present participles like *What is it doing raining?* (Kay & Fillmore 1999: 3).

Nevertheless, even extravagant constructions of the latter type differ from snowclones by their lack of a lexically fixed source, thus not fulfilling the first criterion of our definition. To the best of our knowledge, none of the respective patterns are associated with a specific, culturally salient source construction. As a result, snowclones can well be regarded as a subtype of partially filled constructions (as suggested by Sailer 2013), but we maintain that they are nevertheless distinct from other subtypes of the category and thus deserve to be studied in their own right. Finally, as shown at the bottom of Figure 7, snowclones can also be compared to a range of other extravagant constructions: Kempf & Hartmann (2022), for instance, discuss German pseudo-participles such as *besonnenbrillt* ('be-sunglass-ed'), which look like participles but are not derived from a corresponding verb. Although these examples share with snowclones the fact that they are creative and extravagant, they do not consist of a mix of lexically fixed elements and open slots and are therefore clearly different from the formulaic constructions discussed thus far.

By delineating the category of 'snowclones' in this way, the concept can potentially be used to account for a number of constructions that have been previously discussed on their own terms and without reference to a more general phenomenon. Among these patterns are the 'expressive *much*' or 'sarcastic *much*' construction (Gutzmann & Hendeson 2019, Hilpert & Bourgeois 2020), as in *angry much?*, as well as the 'expressive subordinate clause construction' (Gutzmann & Turgay 2019), as in German *Der Moment wenn man checkt dass man verschlafen hat* ('The moment you realise you've overslept'). Both of these patterns bear striking similarities to snowclones in the way they function as extravagant phrasal

627

templates. And while the extent to which they rely on a lexically fixed source construction deserves further investigation, they form a rather coherent group of constructions tied to specific discourse contexts (especially computer-mediated communication in the case of the expressive subordinate clause construction). As such, they could arguably be reclassified as snowclones.

Finally, we want to address the relationship between snowclones and (internet) memes, another concept that is frequently invoked in this context (e.g. Hill 2018). While we agree that the two concepts are closely related, we also believe that there are good reasons to keep them apart. For one, the concept of 'meme' is ambiguous. In its original conception by Dawkins (1976), it refers to '[a] cultural element or behavioural trait whose transmission and consequent persistence in a population, although occurring by non-genetic means (esp. imitation), is considered as analogous to the inheritance of a gene'.[20] This differs from its modern internet usage, where memes are more narrowly defined as a type of 'image, video, piece of text, etc., typically humorous in nature, that is copied and spread rapidly by internet users, often with slight variations'.[21] Only the latter notion is akin to snowclones. Even in this sense, however, memes are usually associated with multimodal elements, especially pictures and videos. They have been consequently analysed as multimodal constructions that combine image with text (Dancygier & Vandelanotte 2017a, Zenner & Geeraerts 2018). Snowclones, however, are primarily linguistic units that do not require further visual support. Returning to our two main examples, [*the mother of all* X] and [X *BE the new* Y], these patterns are not tied to specific pictures, nor are they necessarily used for humorous effect, as is usually the case with memes. Of course, this is not to say that snowclones and memes do not display some overlap. The examples we gave in Section 3, for instance, include both a multimodal snowclone ([*I* ♥ X]) and a snowclone that typically appears as part of internet memes ([*One does not simply* VP]). Nevertheless, and for the above reasons, we do not believe that snowclones and memes are interchangeable concepts, but rather that each of them has its own role to play in analyses of speakers' creative language (and media) use.

6. CONCLUSION

In this paper, we have investigated the use of snowclones as a particular type of extravagant formulaic construction. We have suggested a definition of snowclones based on three prototypical criteria: the existence of a lexically fixed source construction, the productivity of one or several variable slots, and the presence of 'extravagant' formal and/or functional characteristics. We have illustrated these characteristics with two case studies of frequently occurring snowclones, [*the mother of all* X] and [X *BE the new* Y]. Our results shed light on the functional

---

[20] Oxford English Dictionary (OED), s.v. *meme*, URL www.oed.com/view/Entry/239909 (last checked March 2, 2023)

[21] OED, s.v. *meme*, URL www.oed.com/view/Entry/239909 (last checked March 2, 2023)

profiles of the two constructions, which are particularly productive in certain semantic domains. Moreover, our analyses of typical collexemes and semantic distances provide at least tentative evidence of the extravagant nature of the two patterns, thus expanding upon previous attempts to operationalise extravagance in corpora (Petré 2017, De Wit et al. 2020). Finally, we have contrasted snowclones with other types of idiomatic constructions, arguing that the specific combination of their three defining features sets them apart from these other phenomena. Based on this, we propose that the concept of snowclones, if properly operationalised, forms a useful addition to the inventory of constructionist and phraseological research.

To round off our discussion, we would like to outline a number of open questions that we could not address in this paper but which could be fruitful areas for future research. First, it would be interesting to examine the 'life course' of snowclones and assess whether they follow a similarly short-lived cycle as the one that has been reported for internet memes (Bülow et al. 2018). A related question concerns the factors that determine the 'fate' of snowclones, which either contribute to their continued use and innovation or lead to their eventual disappearance. Second, in this paper we have focused on snowclones in English, and for obvious reasons, it would be worthwhile to take other languages into account as well. The influence of language contact could also be considered in this context. In German, for example, we find equivalents of the two snowclones discussed in the present paper (see e.g. Weber's [2019] study of the German [X *BE the new* Y] construction). This raises the question to what extent languages develop their own independent snowclone inventories and how commonly they adapt snowclones from English, which can probably be considered the *lingua franca* of contemporary pop culture.

Third, it would be worthwhile to investigate whether snowclones are tied to particular registers, genres, and contexts. Starting from Pullum's (2003) original definition, one might expect the natural habitat of snowclones to be journalistic texts. But a considerable proportion of the data analysed for the present paper also comes from computer-mediated communication (CMC), which is often regarded as particularly innovative (see e.g. Bohmann 2016) and as relatively close to spoken language (see e.g. Zitzen & Stein 2004, Soffer 2010). It is an open question to what extent snowclones are tied to specific niches such as journalism or CMC. In addition, it seems conceivable that some snowclones are even more specialised regarding the contexts in which they tend to occur. Perlman (2020), for instance, lists a number of newspaper headlines instantiating the pattern [X *in the time of COVID-19*] (a riff on the novel title *Love in the time of Cholera*), which suggests that this construction may be largely limited to headlines and titles. Last but not least, it would be desirable to study the relationship between snowclones and other extravagant and/or expressive constructions more systematically. In Section 5, we already pointed to some commonalities between snowclones and several constructions (or construction families) that have been discussed as extravagant, but we expect that other similarities, for example with regard to usage contexts and text type distributions, may also play a relevant role.

629

Overall, the goal of this paper has been to show that snowclones are not merely an intellectual curiosity for linguists, but that they constitute a distinct construction type that deserves to be investigated with the help of systematic criteria and quantitative corpus-based tools. Having outlined our approach and methods in detail here, we hope that they will motivate other studies of how language users' creativity and their desire to be noticed shape current linguistic practices.

THE MOTHER OF ALL ACKNOWLEDGMENTS

DATA AVAILABILITY STATEMENT

The supplementary material, including the datasets and scripts used for the present analyses, can be found here: https://osf.io/97t54/.

COMPETING INTERESTS DECLARATION

The authors declare none.

REFERENCES

Baayen, R. Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics: An international handbook*, vol. 2, 899–919. Berlin & New York: De Gruyter.

Baumann, Andreas & Roland Mühlenbernd. 2022. Less of the same: Modeling horror aequi and extravagance as mechanisms of negative frequency dependence in linguistic diversification. In Andrea Ravignani, Rie Asano, Daria Valente, Francesco Ferretti, Stefan Hartmann, Misato Hayashi, Yannick Jadoul, Mauricio Martins, Yohei Oseki, Evelina Daniela Rodrigues, Olga Vasileva & Slawomir Wacewicz (eds.), *The Evolution of Language. Proceedings of the Joint Conference on Language Evolution* (JCoLE), 50–57. Nijmegen: Max Planck Institute for Psycholinguistics.

Bergs, Alexander. 2018. Learn the rules like a pro, so you can break them like an artist (Picasso): Linguistic aberrancy from a constructional perspective. *Zeitschrift für Anglistik und Amerikanistik* 66.3, 277–293.

Bergs, Alexander. 2019. What, if anything, is linguistic creativity? *Gestalt Theory* 41.2, 173–183.

Bohmann, Axel. 2016. Language change because Twitter? Factors motivating innovative uses of *because* across the English-speaking Twittersphere. In Lauren Squires (ed.), *English in computer-mediated communication* (Topics in English Linguistics 93), 149–178. Berlin & New York: De Gruyter.

Booij, Geert. 2002. Constructional idioms, morphology, and the Dutch lexicon. *Journal of Germanic Linguistics* 14.4, 301–327.

Bülow, Lars, Marie-Luis Merten & Michael Johann. 2018. Internet-Memes als Zugang zu multimodalen Konstruktionen. *Zeitschrift für Angewandte Linguistik* 2018.69, 1–32.

Chollet, François & Joseph J. Allaire. 2018. *Deep learning with R.* Shelter Island, NY: Manning Publications Co.

Culicover, Peter W. & Ray Jackendoff. 1999. The view from the periphery: The English comparative correlative. *Linguistic Inquiry* 30.4, 543–571.

630

Cummins, Chris. 2019. *Pragmatics*. Edinburgh: Edinburgh University Press.

Dammerer, Emmanuel Maria. 2007. Buch von der deutschen Snowclonerey. http://emmanuel.dammerer.at/snowclonerey.

Dancygier, Barbara & Lieven Vandelanotte. 2017a. Internet memes as multimodal constructions. *Cognitive Linguistics* 28.3, 565–598.

Dancygier, Barbara & Lieven Vandelanotte. 2017b. Reappraising 'snowclones': Replicability and Construction Grammar. Presented at the 14th International Cognitive Linguistics Conference (ICLC), University of Tartu, Estonia.

Davies, Mark. 2008. The Corpus of Contemporary American English (COCA). https://www.english-corpora.org/coca/.

Davies, Mark. 2010. The Corpus of Historical American English (COHA). https://www.english-corpora.org/coha/.

Dawkins, Richard. 1976. *The selfish gene*. Oxford: Oxford University Press.

De Pascale, Stefano. 2019. *Token-based vector space models as semantic control in lexical lectometry*. Ph.D. dissertation, KU Leuven.

De Wit, Astrid, Peter Petré & Frank Brisard. 2020. Standing out with the progressive. *Journal of Linguistics* 56.3, 479–514.

Desagulier, Guillaume. 2017. *Corpus linguistics and statistics with R: Introduction to quantitative methods in linguistics*. New York: Springer.

Dobrovol'skij, Dimitrij. 2013. Phraseologie und Konstruktionsgrammatik. In Alexander Ziem & Alexander Lasch (eds.), *Konstruktionsgrammatik: Konzepte und Grundlagen gebrauchsbasierter Ansätze* (Germanistische Arbeitshefte 44), 111–130. Berlin & New York: De Gruyter.

Dunning, Ted. 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics* 19.1, 69–74.

Eitelmann, Matthias & Dagmar Haumann (eds.). 2022. Extravagant morphology: *Studies in rule-bending, pattern-extending and theory-challenging morphology* (Studies in Language Companion Series 223). Amsterdam & Philadelphia: John Benjamins.

Fauconnier, Gilles & Mark Turner. 2002. *The way we think: Conceptual blending and the mind's hidden complexities*. New York: Basic Books.

Ferguson, Kevin L. 2019. *Pop goes the decade: The nineties*. Santa Barbara, CA: Greenwood.

Fillmore, Charles J., Paul Kay & Mary Catherine O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language* 64.3, 501–538.

Finkbeiner, Rita. 2019. *Tschüssikowski* und *Bis später, Attentäter*: Zur Bedeutung von expressive Verabschiedungen. In Franz-Josef d'Avis & Rita Finkbeiner (eds.), *Expressivität im Deutschen* (Reihe Germanistische Linguistik 318), 341–369. Berlin & Boston: De Gruyter.

Flach, Susanne. 2021. *collostructions: An R implementation for the family of collostructional methods*. R package version v.0.2.0. https://sfla.ch/collostructions/.

Gaeta, Livio & Davide Ricca. 2006. Productivity in Italian word-formation. *Linguistics* 44.1, 57–89.

Glucksberg, Sam & Boaz Keysar. 1993. How metaphors work. In Andrew Ortony (ed.), *Metaphor and Thought,* 2nd edn., 401–424. Cambridge: Cambridge University Press.

Grady, Joseph E. 1999. A typology of motivation for conceptual metaphor: Correlation vs. resemblance. In Raymond W. Gibbs & Gerard J. Steen (eds.), *Metaphor in cognitive linguistics: Selected papers from the 5th International Cognitive Linguistics Conference, Amsterdam, 1997* (Current Issues in Linguistic Theory 175), 79–100. Amsterdam & Philadelphia: John Benjamins.

Gries, Stefan Th. 2019. 15 years of collostructions: Some long overdue additions/corrections (to/of actually all sorts of corpus-linguistics measures). *International Journal of Corpus Linguistics* 24.3, 385–412.

Gutzmann, Daniel & Robert Henderson. 2019. Expressive updates, much? *Language* 95.1, 107–135.

Gutzmann, Daniel & Katharina Turgay. 2019. Der Moment, wenn dir klar wird, dass es expressive Nebensätze gibt. Zur Syntax und Semantik der expressiven Nebensatzkonstruktion. In Franz-Josef d'Avis & Rita Finkbeiner (eds.), *Expressivität im Deutschen* (Reihe Germanistische Linguistik 318), 261–291. Berlin & Boston: De Gruyter.

Hartmann, Stefan. 2014. Constructing a schema: Word-class changing morphology in a usage-based perspective. *Yearbook of the German Cognitive Linguistics Association* 2.1, 235–252.

Hartmann, Stefan. 2018. Derivational morphology in flux: A case study of word-formation change in German. *Cognitive Linguistics* 29.1, 77–119.

Haspelmath, Martin. 1999. Why is grammaticalization irreversible? *Linguistics* 37.6, 1043–1068.

631

Hill, Ian E. J. 2018. Memes, munitions, and collective copia: The durability of the perpetual peace weapons snowclone. *Quarterly Journal of Speech* 104.4, 422–443.

Hilpert, Martin & Samuel Bourgeois. 2020. Intersubjectification in constructional change: From confrontation to solidarity in the *sarcastic much?* construction. *Constructions and Frames* 12.1, 96–120.

Hilpert, Martin & Florent Perek. 2015. Meaning change in a petri dish: Constructions, semantic vector spaces, and motion charts. *Linguistics Vanguard* 1.1, 339–350.

Hilpert, Martin & Florent Perek. 2022. You don't get to see that every day: On the development of permissive *get. Constructions and Frames* 14.1, 13–40.

Howarth, Peter A. 1996. *Phraseology in English academic writing: Some implications for language learning and dictionary making.* Tübingen: Niemeyer.

Jurafsky, Daniel & James H. Martin. 2023. *Speech and natural language processing: An introduction to Natural Language Processing, computational linguistics, and speech recognition* (3rd edn. draft). https://web.stanford.edu/~jurafsky/slp3/ed3book_jan72023.pdf.

Kay, Paul & Charles J. Fillmore. 1999. Grammatical constructions and linguistic generalizations: The *What's X doing Y?* construction. *Language* 75.1, 1–33.

Keller, Rudi. 1994. *Sprachwandel: Von der unsichtbaren Hand in der Sprache.* Tübingen & Basel: Francke.

Kempf, Luise & Stefan Hartmann. 2022. What's extravagant about *be-sandal-ed feet*? Morphology, semantics and pragmatics of German pseudo-participles. In Matthias Eitelmann & Dagmar Haumann (eds.), *Extravagant morphology. Studies in rule-bending, pattern-extending and theory-challenging morphology*, 19–50. Amsterdam & Philadelphia: John Benjamins.

Krijthe, Jesse H. 2015. *Rtsne: t-distributed stochastic neighbor embedding using Barnes-Hut implementation.* R package version 0.16. https://github.com/jkrijthe/Rtsne.

Lass, Roger. 2015. Lineage and the constructive imagination: The birth of historical linguistics. In Claire Bowern & Bethwyn Evans (eds.), *The Routledge handbook of historical linguistics*, 45–63. New York: Routledge.

Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis.* Amsterdam & Philadelphia: John Benjamins.

Levshina, Natalia & Kris Heylen. 2014. A radically data-driven Construction Grammar: Experiments with Dutch causative constructions. In Ronny Boogaart, Timothy Colleman & Gijsbert Rutten (eds.), *Extending the scope of construction grammar* (Cognitive Linguistics Research 54), 17–46. Berlin & New York: De Gruyter.

Levy, Omer. 2022. Word representation. In Ruslan Mitkov (ed.), *The Oxford handbook of computational linguistics,* 2nd edn., 334–358. Oxford: Oxford University Press.

Liberman, Mark. 2006. The proper treatment of snowclones in ordinary English. http://itre.cis.upenn.edu/~myl/languagelog/archives/002806.html.

van der Maaten, Laurens. 2014. Accelerating t-SNE using tree-based algorithms. *Journal of Machine Learning Research* 15.93, 3221–3245.

van der Maaten, Laurens & Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9.86, 2579–2605.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado & Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In *NIPS'13: Proceedings of the 26th International Conference on Neural Information Processing Systems* (NIPS) 2, 3111–3119. Red Hook, NY: Curran Associates Inc.

Nguyen, Kim Anh, Sabine Schulte im Walde & Ngoc Thang Vu. 2016. Integrating distributional lexical contrast into word embeddings for antonym-synonym distinction. In Katrin Erk & Noah A. Smith (eds.), *Proceedings of the 54th annual meeting of the association for computational linguistics, volume 2: Short papers (ACL)*, 454–459. Berlin: Association for Computational Linguistics.

Nguyen, Kim Anh, Sabine Schulte im Walde & Ngoc Thang Vu. 2017. Distinguishing antonyms and synonyms in a ppttern-based neural network. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, volume 1: Long papers* (ACL), 76–85. Valencia, Spain: Association for Computational Linguistics.

O'Connor, Erin. 2007. The snowclones database. https://snowclones.org/.

Perek, Florent. 2016. Using distributional semantics to study syntactic productivity in diachrony: A case study. *Linguistics* 54.1, 149–188.

Perek, Florent & Martin Hilpert. 2017. A distributional semantic approach to the periodization of change in the productivity of constructions. *International Journal of Corpus Linguistics* 22.4, 490–520.

Perlman, Merrill. 2020. Snowclones in the time of pandemic. https://www.cjr.org/language_corner/pandemic-covid-19-time-of-cholera-marquez.php.

Petré, Peter. 2017. The extravagant progressive: An experimental corpus study on the history of emphatic [be V*ing*]. *English Language and Linguistics* 21.2, 227–250.

Pullum, Geoffrey K. 2003a. Bleached conditionals. http://itre.cis.upenn.edu/~myl/languagelog/archives/000049.html.

Pullum, Geoffrey K. 2003b. Phrases for lazy writers in kit form. http://itre.cis.upenn.edu/~myl/languagelog/archives/000061.html.

Pullum, Geoffrey K. 2004. Snowclones: Lexicographical dating to the second. http://itre.cis.upenn.edu/~myl/languagelog/archives/000350.html.

R Core Team. 2020. *R: A language and environment for statistical computing*. Vienna, Austria. https://www.R-project.org/.

Sailer, Manfred. 2013. Idiom and phraseology. *Oxford Bibliographies*. https://oxfordbibliographies.com/view/document/obo-9780199772810/obo-9780199772810-0137.xml.

Sampson, Geoffrey. 2016. Two ideas of creativity. In Martin Hinton (ed.), *Evidence, experiment and argument in linguistics and philosophy of language* (Studies in Philosophy of Language and Linguistics 3), 15–26. Bern: Peter Lang.

Schmid, Hans-Jörg. 2020. *The dynamics of the linguistic system: Usage, conventionalization, and entrenchment*. Oxford: Oxford University Press.

Schmidt, Ben and Jian Li. 2022. *wordVectors: Tools for creating and analyzing vector-space models of texts*. R package version 2.0. http://github.com/bmschmidt/wordVectors.

Schneider, Ulrike. 2020. ΔP as a measure of collocation strength: Considerations based on analyses of hesitation placement in spontaneous speech. *Corpus Linguistics and Linguistic Theory* 16.2, 249–274.

Schulze, Florian. 2013. Noch so 'ne Phrase, Faust auf die Nase! Eine phraseologische Untersuchung des Nerv-Sprech. *Linguistik Online* 60.3, 59–78.

Schäfer, Roland. 2015. Processing and querying large web corpora with the COW14 architecture. In Piotr Baski, Hanno Biber, Evelyn Breiteneder, Marc Kupietz, Harald Lüngen & Andreas Witt (eds.), *Proceedings of the 3rd Workshop on Challenges in the Management of Large Corpora* (CMLC-3), 28–34. Mannheim: Institut für Deutsche Sprache.

Schäfer, Roland & Felix Bildhauer. 2012. Building large corpora from the web using a new efficient tool chain. In Nicoletta Calzolari, Khalid Choukri, Terry Declerck, Mehmet Uur Doan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation* (LREC'12), 486–493. Istanbul: European Language Resources Association.

Slowikowski, Kamil. 2021. *ggrepel: Automatically position non-overlapping text labels with 'ggplot2'*. R package version 0.9.1. https://CRAN.R-project.org/package=ggrepel.

Smirnova, Elena. 2021. Deutsche Partizipialkomposita aus diachroner Perspektive. In Christine Ganslmayer & Christian Schwarz (eds.), *Historische Wortbildung: Theorie – Methode – Perspektiven*, 383–415. Hildesheim: Olms.

Soffer, Oren. 2010. Silent orality: Toward a conceptualization of the digital oral features in CMC and SMS texts. *Communication Theory* 20.4, 387–404.

Stefanowitsch, Anatol. 2013. Collostructional analysis. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford handbook of Construction Grammar*, 290–306. Oxford: Oxford University Press.

Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8.2, 209–243.

Stefanowitsch, Anatol & Stefan Th Gries. 2005. Covarying collexemes. *Corpus Linguistics and Linguistic Theory* 1.1, 1–43.

Tizón Couto, David. 2021. The rise of COVID 'snowclones': The mother of all linguistic phrases. https://theconversation.com/the-rise-of-covid-snowclones-the-mother-of-all-linguistic-phrases-167580.

Traugott, Elizabeth Closs. 2014. Towards a constructional framework for studying language change. *Cognitive Linguistic Studies* 1.1, 3–21.

Traugott, Elizabeth Closs & Graeme Trousdale. 2013. *Constructionalization and constructional changes*. Oxford: Oxford University Press.

Traugott, Elizabeth Closs & Graeme Trousdale. 2014. Contentful constructionalization. *Journal of Historical Linguistics* 4.2, 256–283.

633

Ungerer, Tobias & Stefan Hartmann. 2020. Delineating extravagance: Assessing speakers' perceptions of imaginative constructional patterns. *Belgian Journal of Linguistics* 34, 345–356.

Weber, Tilo. 2019. Das Wechselspiel zwischen Lexikon und Diskurs: So genannte Antisprichwörter und die *X ist das neue Y*-Konstruktion. *Linguistik Online* 96.3, 133–156.

Wheeler, Eric S. 2005. Multidimensional scaling for linguists. In Reinhard Köhler, Gabriel Altmann & Rajmund G. Piotrowski (eds.), *Quantitative linguistics: An international handbook* (Handbooks of Linguistics and Communication Science 27), 548–553. Berlin & New York: De Gruyter.

Zeldes, Amir. 2013. *Productivity in argument selection: From morphology to syntax*. Berlin & Boston: De Gruyter.

Zenner, Eline & Dirk Geeraerts. 2018. One does not simply process memes: Image macros as multimodal constructions. In Esme Winter-Froemel & Verena Thaler (eds.), *Cultures and traditions of wordplay and wordplay research* (The Dynamics of Wordplay 6), 167–194. Berlin & Boston: De Gruyter.

Zimmer, Ben. 2007. Phrasal patterns 2: Electric boogaloo. https://blog.oup.com/2007/08/patterns/.

Zimmer, Benjamin. 2006. On the trail of "the new black" (and "the navy blue"). http://itre.cis.upenn.edu/~myl/languagelog/archives/003981.html.

Zitzen, Michaela & Dieter Stein. 2004. Chat and conversation: a case of transmedial stability? *Linguistics* 42.5.

Zwicky, Arnold. 2006a. Fully awesome! http://itre.cis.upenn.edu/%7emyl/languagelog/archives/003764.html.

Zwicky, Arnold. 2006b. Snowclone mountain? http://itre.cis.upenn.edu/~myl/languagelog/archives/002924.html.

Authors' addresses: *(Hartmann)*
University of Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany
hartmast@hhu.de

(Ungerer)
Concordia University, 7141 Sherbrooke Street West, Montreal, QC, H4B 1R6, Canada
tobias.ungerer@concordia.ca