

ojs.uv.es/index.php/qfilologia/index

Rebut: 01.06.2024. **Acceptat:** 05.07.2024

Per a citar aquest article: Ungerer, Tobias & Hartmann, Stefan. 2024. "Contrastive is the new black: A cross-linguistic study of a 'snowclone' in English, German, and Spanish". *Quaderns de Filologia: Estudis Lingüístics* XXIX: 217-235.

doi: 10.7203/QF.29.28712



Contrastive is the new black: A cross-linguistic study of a “snowclone” in English, German, and Spanish

Contrastivo es el nuevo negro: un estudio translingüístico de un “snowclone” en inglés, alemán y español

TOBIAS UNGERER
University of Toronto
tobias.ungerer@utoronto.ca

STEFAN HARTMANN
Heinrich Heine University Düsseldorf
hartmast@hhu.de

Abstract: Recent constructionist research has seen increased interest in language contact and contact-induced change. In this paper, we present a contrastive corpus-based analysis of the “snowclone” [X BE the new Y] and its equivalents in German and Spanish. Our results show that the construction is most frequent in English but productive in all three languages. Collostruational analysis techniques reveal parallels among the lexemes that the pattern combines with in each language, but also differences in the degree of semantic variability. Together, the results reflect the role of constructional borrowing in the emergence of the German and Spanish constructions, which can be modelled via links in the mental network of multilingual speakers.

Keywords: construction grammar; language contact; phraseology; constructional borrowing; snowclones.

Resumen: La investigación construccionista reciente ha experimentado un creciente interés en el contacto lingüístico y el cambio inducido por el contacto. En este artículo presentamos un análisis contrastivo basado en corpus del “snowclone” [X BE the new Y] y sus equivalentes en alemán y español. Nuestros resultados muestran que la construcción es más frecuente en inglés, pero productiva en los tres idiomas. Las técnicas de análisis colostruacional revelan paralelismos entre los lexemas con los que se combina el patrón en cada lengua, pero también diferencias en el grado de variabilidad semántica. En resumen, los resultados reflejan el papel del préstamo construccionista en el surgimiento de las construcciones alemana y española, que pueden modelarse a través de vínculos en la red mental de hablantes multilingües.

Palabras clave: gramática de construcciones; contacto lingüístico; fraseología; préstamo constructivo; *snowclones*.

1. Introduction

While early work in construction grammar remained largely focused on monolingual research, the last 15 years have seen a surge of interest in how language contact can be modelled from a constructionist perspective (e. g., Pietsch, 2010; Van de Velde & Zenner, 2010; Wasserscheidt, 2014; Coleman, 2016; Zenner, Heylen & Van de Velde, 2018; De Pascale, Pijpops, Van de Velde & Zenner, 2022). Case studies of contact-induced change have been conducted in a range of languages, including Afrikaans, American Sign Language, Dutch, Malayalam, Swedish, and Welsh, as well as specific language varieties, such as Cajun French, Hiberno English, and Texas German (see, e. g., the edited volumes by Boas & Höder, 2018, 2021). Moreover, theoretical approaches have been proposed, most notably Höder's (2012, 2018) diasystematic construction grammar, which account for contact phenomena using constructionist concepts, such as form-meaning pairings at varying levels of schematicity and their organization in multilingual constructional networks.

In this paper, we aim to contribute to this growing research trend by analyzing another putative case of language contact: the extension of the pattern [X *BE the new* Y] from English to German and Spanish. This construction instantiates a type of semi-fixed idiomatic formulae known as “snowclones” (Traugott & Trousdale, 2013; Hartmann & Ungerer, 2024), which are originally derived from a more lexically fixed source (in this case, fashion slogans such as *pink is the new black*), but which have developed open slots that can be filled by variable lexical items. While the construction is particularly well documented in English (Hartmann & Ungerer, 2024), as illustrated in (1), it also occurs in German (Weber, 2019), as shown in (2), and it is further attested in Spanish; compare (3) (all examples are taken from the COW web corpora; see section 3.1).

- (1) a. *This year, grey is the new black*
 b. *Thursday is the new Friday. Sleep is overrated.*
- (2) a. *Die Sechziger sind die neuen Achtziger! Zumindest dann, wenn es um Sonnenbrillen-Mode geht*
 ‘The sixties are the new eighties! At least when it comes to sunglasses fashion’
 b. *Teilen ist das neue Besitzen. Wer Dinge nicht mehr braucht, gibt sie weiter*

‘Sharing is the new owning. If you no longer need things, you pass them on’.

(3) a. *La ironía es el nuevo negro: a los humanos nos gusta el humor*

‘Irony is the new black: we humans like humor’

b. *Hoy las drogas son el nuevo alcohol, y se está cometiendo el mismo error de siempre.*

‘Nowadays, drugs are the new alcohol, and the same mistake as always is being made’.

Our goal in this study is twofold: first, we provide a comparative corpus-based analysis of how [X BE the new Y] is used in English, German, and Spanish, focusing in particular on the productivity of the construction and the semantic variability of its slot fillers. In doing so, we go beyond previous constructionist work, which has usually focused on comparing only two languages, and explore how the characteristics of each of our languages affect speakers’ use of the pattern. Second, we will discuss how the status of the construction in the three languages can be modelled from a constructionist perspective. This includes both the question of whether the German and Spanish patterns may be the result of “constructional borrowing” (Coleman, 2016) from English, and how the three language-specific constructions relate to one another in the mental network of multilingual speakers.

2. Background

2.1 [X BE the new Y] across languages

The term “snowclone”, which was proposed by authors of the linguistic blog *Language Log* (Pullum, 2004), refers to a specific type of semi-fixed formulaic construction (Traugott & Trousdale, 2013: 183-186; Hill, 2018). Hartmann & Ungerer (2024: 605) define snowclones via three criteria: *i*) they are based on a culturally salient lexically fixed source; *ii*) they are extended to new instances via lexical substitution in one or several variable slots; *iii*) they display distinctive or “extravagant” (Haspelmath, 1999) formal and/or functional characteristics that function as markers of linguistic innovation. A typical example is [*the mother of all X*], which, in its present-day usage, goes back to a quote by

Saddam Hussein (*the mother of all battles*). The pattern contains an open noun slot and expresses a striking hyperbolic meaning ('an extreme example of X').

[X BE *the new* Y], illustrated above in (1), also matches the above snowclone definition if the first criterion is somewhat loosely interpreted, given that the pattern is usually traced back to an amalgam of 1970s/80s fashion slogans rather than to a single lexical source. The construction, which is included in major English dictionaries, is "used to say that something is now more popular or fashionable than the thing that it replaces"¹. Hartmann & Ungerer (2024) provide a corpus analysis of around 3,850 instances of the pattern, showing that the construction is highly productive, and that its slots can be filled by both nouns and adjectives from various semantic domains. Nevertheless, the results also reveal semantic preferences for certain domains, such as colors, group membership terms, and internet-related concepts, a point that we will return to in our analysis in section 3.

Regarding the occurrence of [X BE *the new* Y] in other languages, the only empirical study to date seems to be Weber's (2019) analysis of the pattern in German, which is based on around 520 tokens from the German Reference Corpus (DeReKo). While normalized corpus frequencies are not reported, the sample size suggests that the snowclone is productive in German, even though perhaps to a lesser degree than in English. This is also in line with the fact that an example of the pattern is included in the German *Duden* dictionary (Duden, "neu", sense 5a), even though the construction does not have its separate (sub-)entry. Nevertheless, given that Weber's (2019) analysis focuses on qualitative orthographic and syntactic features of the snowclone, a quantitative assessment of its productivity and degree of semantic variability in German is still lacking.

For other languages, including Spanish, no comparable studies of [X BE *the new* Y] have been conducted. Based on an informal search of web forums and consultations with Spanish native speakers, our initial expectation was that the snowclone may be less productive in Spanish, and that some speakers may even perceive it as unacceptable.

As far as the origin of the German and (putative) Spanish [X BE *the new* Y] patterns is concerned, the widespread assumption seems to be that they are the result of transfer from English, even though the evidence for this remains tentative. Weber (2019: 146, fn. 6) refers to "a number of indications" that the German construction is influenced by English, for example the fact that

¹ <https://dictionary.cambridge.org/dictionary/english/be-the-new> [Access 02/05/2024].

many instances of the pattern have well-established translation equivalents in English. Rühle (2010), in a newspaper comment on the topic, claims that the expression entered the German language around 2000, i. e., well after it became established in English. For Spanish, even less information is available, but a similar case can plausibly be made, especially given anecdotal evidence that many internet forum users perceive the pattern as a loan from English.

Based on the working hypothesis that [X *BE the new* Y] is subject to cross-linguistic transfer from English, we will next discuss how such instances of language contact can be accounted for from a constructionist perspective.

2.2 Phrasal borrowings from a constructionist perspective

While research on language contact at the lexical level was, for a long time, focused on single words (Onysko 2007), more recent work has increasingly targeted multi-word units, addressing them either from a phraseological perspective (e. g., Furiassi, Pulcini & Rodríguez González, 2012; Fiedler, 2017; Andersen, 2022) or a constructionist angle (e. g., Pietsch, 2010; Van de Velde & Zenner, 2010; Wasserscheidt, 2014; Coleman, 2016; Zenner et al., 2018; Boas & Höder, 2018, 2021; De Pascale et al., 2022). Some of these studies have focused on lexically fixed patterns such as German *die Extrameile gehen* (borrowed from English *(to) go the extra mile*; Fiedler, 2017), but others have investigated partially flexible patterns with open slots.

Coleman (2016), for example, discusses the close parallels between the English ‘time’-*away* construction in (4a) and an emerging Dutch construction with the particle *weg* ‘away’, illustrated in (4b) (examples from Coleman, 2016: 94, 99).

- (4) a. *Bill slept the afternoon away.*
 b. *Daar hebben we de avond weggedronken en gedanst.*
 ‘There we drank and danced the evening away’.

Given that the Dutch pattern, as Coleman argues, is likely the result of transfer from English, he regards it as an instance of “constructional borrowing”. This concept is used to highlight that language users transfer partially schematic form-meaning pairs, rather than fully lexicalized phrases, from one language to another.

In addition, Coleman discusses how the relationship between the English and the Dutch construction can be accounted for within Höder's (2012, 2018) diasystematic construction grammar. In this framework, language contact is regarded as a ubiquitous phenomenon that occurs in multilingual speakers' minds. Crucially, these speakers are assumed to generalize over similar constructions in their individual languages (called "idioconstructions") and form more abstract language-independent representations (called "diaconstructions"). Applying these concepts to his case study, Coleman (2016: 103) argues that speakers may represent a diaconstruction that captures the shared structure and semantics of the English and Dutch 'time'-*away/weg* constructions, and which merely leaves the language-dependent particle underspecified.

Building on this previous work, we will examine how [X *BE the new* Y], as a frequent English snowclone, is used in German and Spanish. In particular, we will address the following three questions:

- How productive and semantically variable is [X *BE the new* Y] in the three languages?
- Are there signs that the German and Spanish patterns are the result of constructional borrowing from English?
- How can this contact situation be modelled from the perspective of usage-based construction grammar?

3. Corpus study

3.1 *Material and methods*

Our corpus study draws on the COW family of corpora (Schäfer & Bildhauer, 2012; Schäfer, 2015). COW is a suite of web corpora covering different languages, including the three object languages of the present study, English, German, and Spanish. The data were crawled in the early 2010s. Compared to other multi-language web corpora such as the TenTen family of corpora (Jakubíček, Kilgarriff, Kovář, Rychlý & Suchomel, 2013), which is updated with new crawls on a regular basis, the COW data have the disadvantage that they may not give the most up-to-date picture of the development of a pattern. They do, however, have the significant advantage that they are available free of charge and that large parts of the full data as well as aggregated data such as word frequency lists can be easily accessed, which is an important prereq-

uisite for the kinds of quantitative approaches employed here. We searched for all instances of [X *BE the new* Y] in the respective languages, taking all inflectional forms of the verb and, where applicable, the adjective into account.

Table 1 provides a summary of the corpus sizes, the number of hits, and the number of true positives after manual checking. For the English data, we draw on the dataset from Hartmann & Ungerer (2024), which is based on a sample of 5,000 instances, yielding 3,865 true positives². Since the proportion of false positives is much larger in the German and the Spanish corpora, we checked the entire datasets, yielding 1,080 true positives for German and 267 for Spanish.

<i>Corpus</i>	<i>Corpus size</i>	<i>Hits</i>	<i>True positives</i>
ENCOW16A (English)	16,502,849,469	5,000 (sample out of 23,992)	3,865
ESCOW16A (Spanish)	7,131,363,847	4,811	267
DECOW16B (German)	19,833,893,137	55,320	1,080

Table 1. Overview of the datasets used for the present study

In the remainder of this section, we adopt some of Hartmann & Ungerer’s (2024) methods for investigating English snowclones and apply them to our German and Spanish datasets.

First, we use simple frequency measures to gauge the overall productivity of the pattern in each language. Besides token frequencies (i. e., the number of instances of a pattern overall), we analyze type frequencies (i. e., the number of *different* instances) and the number of hapax legomena (i. e., instances that are only attested a single time). Type frequencies can be regarded as an indicator of “realized productivity” (Baayen, 2009), expressing the degree of lexical variation attested in the data. Meanwhile, hapax frequencies, when divided by the number of tokens, are used as a measure of “potential productivity” (Baayen, 2009), capturing the likelihood that a pattern may give rise to new coinages. Baayen (2009: 902) compares this to companies: A company may have a large share of the market (i. e., high type frequency), but at some point the market may be saturated because there are no buyers left. In con-

² Hartmann & Ungerer (2024) report 3,848 because 17 hits with unclear concept annotation were excluded from the analysis.

trast, a high hapax-token ratio indicates that the market is not yet saturated, and that further expansions of the company can be expected in the future.

As a second step, we investigate the semantic constraints and preferences of [X BE the new Y] in each language, relying on collocation analysis. Specifically, we use simple collexeme analysis (Stefanowitsch & Gries, 2003) to examine which words are strongly attracted to the X and the Y slot of the snowclone, and covarying collexeme analysis (Stefanowitsch & Gries, 2005) to test what combinations of X and Y elements are particularly typical. Both methods involve the computation of an association measure over a contingency table, which contains the frequencies of a word and all the other words in and outside of the construction in the case of simple collexeme analysis, as well as the frequencies of a word and all the other words in each slot of the construction in the case of covarying collexeme analysis. For calculating the collexeme strengths, we used Flach's (2021) package *collocations* for R (R Core Team 2024). We report the log-likelihood ratio G^2 as an association measure.

As a final step, we conduct a contrastive collocation analysis (Gilquin, 2015). We automatically translated the lemmas in the X and Y slots of the German and Spanish datasets to English using ChatGPT 4.0 and used multiple distinctive collexeme analysis across the lemma pairs in the three languages to detect which items are significantly attracted to each of the three languages. For this analysis, we use the *t*-score (Church, Gale, Hanks & Hindle, 1991) as the association measure of our choice, following, e. g., Paquot & Bestgen (2009; also see Gries, 2024, for discussion).

3.2 Results

3.2.1 Frequency measures

Table 2 summarizes the token, type, and hapax frequencies of [X BE the new Y]; for the latter two, separate counts are provided for the X and the Y slot. Starting with token frequencies, the results show that the construction is clearly most frequent in English, especially given that we only analyzed a subset of our query hits, which means that the frequency of the pattern in the entire corpus should be approximately 4 to 5 times higher (corresponding to an estimated normalized frequency of 1.12 per million). In German, the construction is less frequent than in English; note that the two corpora have similar sizes (see section 3.1). Meanwhile, the German corpus is almost three

times as large as the Spanish one. Adjusting for this difference, the normalized token frequency is only somewhat higher for the German construction (0.054 per million) than in Spanish (0.037 per million).

	Tokens	Types		Hapax legomena	
		X	Y	X	Y
English	3,865	2,234 (0.58)	1,754 (0.45)	1,809 (0.47)	1,378 (0.36)
German	1,080	573 (0.53)	522 (0.48)	443 (0.41)	400 (0.37)
Spanish	266	190 (0.71)	168 (0.63)	155 (0.58)	136 (0.51)

Table 2. Token, type, and hapax frequencies in each language; type-token ratios and hapax-to-token ratios are provided in brackets

Relative to these token frequencies, the type frequencies and hapax counts are high in all three languages. As shown by the type-token ratios provided in brackets in table 2, the number of types is between 45 % and 71 % as high as the number of tokens, with the exact values depending on the slot and language. Moreover, the hapax counts indicate that many of these types only occur once, making up 36 % to 58 % of tokens. For Spanish, the ratios appear to be higher than for the other languages, but this is likely an artifact of the difference in corpus size (Gaeta & Ricca, 2006; Hartmann, 2018), so no further conclusion should be drawn from this fact. Meanwhile, an interesting observation is that, within each language, the type and hapax frequencies are somewhat higher in the X slot than in the Y slot. As already suggested by Hartmann & Ungerer (2024: 618-619) for the English data, this result may be driven by the fact that one particular Y item (*black/schwarz/negro*) occurs particularly frequently (see the results in section 3.2.2) and thus reduces the amount of variation in the Y slot.

3.2.2 Collostructional analyses

Starting with the results of the simple collexeme analyses, table 3 shows the top 10 most strongly attracted collexemes for the X slot and the Y slot in each language³. The full collexeme tables, along with additional association meas-

³ German *datum* and Spanish *dato* are only attested in their plural forms, where they mean ‘data’.

ures, are available in the online supplementary material (see section “Data availability” at the end of the paper).

Language	X slot			Y slot		
	Collexeme	Freq in cxn	G ²	Collexeme	Freq in cxn	G ²
English	<i>green</i>	91	500.12	<i>black</i>	622	5,503.80
	<i>pink</i>	40	298.4	<i>sexy</i>	32	282.77
	<i>black</i>	67	282.79	<i>oil</i>	57	278.08
	<i>small</i>	77	250.03	<i>big</i>	75	269.42
	<i>white</i>	55	203.29	<i>nazi</i>	26	200.97
	<i>grey</i>	27	175.89	<i>pink</i>	27	182.49
	<i>blog</i>	38	161.37	<i>jews</i>	29	156.7
	<i>brown</i>	32	145.39	<i>cupcake</i>	14	151.64
	<i>facebook</i>	27	138.50	<i>kkk</i>	12	145.88
	<i>data</i>	51	137.40	<i>objectivity</i>	14	131.64
German	<i>teilen</i> ‘share’	39	386.84	<i>schwarz</i> ‘black’	135	1283.44
	<i>grün</i> ‘green’	44	343.53	<i>haben</i> ‘have’	27	323.35
	<i>blau</i> ‘blue’	34	275.57	<i>pink</i>	23	272.28
	<i>grau</i> ‘gray’	25	223.53	<i>öl</i> ‘oil’	23	190.26
	<i>datum</i> ‘data’	29	165.02	<i>jude</i> ‘jew’	18	143.29
	<i>weiß</i> ‘white’	26	161.06	<i>grün</i> ‘green’	21	132.82
	<i>uncool</i>	11	155.09	<i>cool</i>	17	120.98
	<i>orange</i>	14	134.65	<i>punk</i>	11	109.03
	<i>selbstjustiz</i> ‘vigilantism’	4	53.22	<i>indie</i>	8	105.57
	<i>silber</i> ‘silver’	7	52.77	<i>gold</i>	14	102.41

Language	X slot			Y slot		
	Collexeme	Freq in cxn	G ²	Collexeme	Freq in cxn	G ²
Spanish	<i>dato</i> ‘data’	11	71.99	<i>negro</i> ‘black’	28	291.62
	<i>internet</i>	8	55.38	<i>petróleo</i> ‘oil’	10	116.09
	<i>gris</i> ‘gray’	5	52.31	<i>publicidad</i> ‘advertising’	8	75.43
	<i>transparencia</i> ‘transparency’	5	50.07	<i>objetividad</i> ‘objectivity’	5	66.18
	<i>ecologismo</i> ‘environmentalism’	3	49.09	<i>comunismo</i> ‘communism’	4	52.29
	<i>neoliberalismo</i> ‘neoliberalism’	3	39.8	<i>esclavitud</i> ‘slavery’	4	49.64
	<i>40</i>	5	37.04	<i>aristocracia</i> ‘aristocracy’	3	42.33
	<i>preadolescente</i> ‘preadolescent’	2	33.45	<i>40</i>	5	36.97
	<i>azul</i> ‘blue’	4	32.53	<i>rock</i>	4	36.64
	<i>perchero-lámpara</i> ‘coat rack-lamp’	1	31.85	<i>zoofilia</i> ‘zoophilia’	1	31.84

Table 3. Results of the simple collexeme analyses for the X slot and the Y slot

The results of the simple collexeme analysis show some interesting commonalities among the three languages, but also a few relevant differences. One aspect that the attracted collexemes in all three languages have in common is that colors feature prominently, which points to the use of [X BE the new Y] as a fashion slogan, which is often seen as the original ‘niche’ of the snowclone (see Hartmann & Ungerer, 2024). Also, *black* and its translation equivalents feature most prominently among the attracted collexemes in the Y slot, thus accounting for 16.1 % (English), 12.5 % (German), and 10.5 % (Spanish) of all tokens. This provides evidence for the assumption of a sub-construction [X BE the new *black*] that is so frequent across all three languages that it can be considered a pattern in its own right.

Language	X slot collexeme	Y slot collexeme	Freq in cxn	G ²
English	<i>small</i>	<i>big</i>	75	720.65
	<i>data</i>	<i>oil</i>	37	283.66
	<i>blog</i>	<i>resume</i>	20	198.98
	<i>transparency</i>	<i>objectivity</i>	14	163.3
	<i>muslims</i>	<i>jews</i>	15	139.99
	<i>green</i>	<i>red</i>	23	138.36
	<i>old</i>	<i>new</i>	13	126.99
	<i>strong</i>	<i>skinny</i>	10	118.28
	<i>quiet</i>	<i>loud</i>	8	114.8
	<i>nra</i>	<i>kkk</i>	9	113.53
German	<i>teilen</i> 'sharing'	<i>haben</i> 'having'	27	204.37
	<i>datum</i> 'data'	<i>öl</i> 'oil'	21	159.36
	<i>klein</i> 'small'	<i>groß</i> 'big'	10	113.55
	<i>uncool</i>	<i>cool</i>	11	100.72
	<i>langsam</i> 'slow'	<i>schnell</i> 'fast'	6	74.28
	<i>vernunft</i> 'reason'	<i>punk</i>	7	70.08
	<i>frau</i> 'woman'	<i>mann</i> 'man'	6	68.54
	<i>herbst</i> 'autumn'	<i>sommer</i> 'summer'	6	68.54
	<i>grün</i> 'green'	<i>pink</i>	14	64.18
	<i>grüne</i> 'green (party)'	<i>fdp</i> 'fdp' (liberal party)	5	58.32
	Spanish	<i>transparencia</i> 'transparency'	<i>objetividad</i> 'objectivity'	5
<i>dato</i> 'data'		<i>petróleo</i> 'oil'	8	48.97
30		20	4	31.52
40		30	4	31.52
60		40	3	26.15
<i>preadolescente</i> 'preadolescent'		<i>adolescente</i> 'adolescent'	2	23.55
<i>cuervo</i> 'body'		<i>contraseña</i> 'password'	2	23.55
<i>pueblo</i> 'people'		<i>gobierno</i> 'government'	2	23.55
<i>azúcar</i> 'sugar'		<i>tabaco</i> 'tobacco'	2	23.55
<i>gris</i> 'gray'		<i>negro</i> 'black'	5	23.37

Table 4. Results of the covarying collexeme analysis

The concept domains of the top 10 attracted collexemes also show some interesting overlaps. In addition to colors, we find terms referring to political

or religious groups and ideologies. A characteristic that sets apart the Spanish data from the German and English ones is that, apart from color terms, no adjectives are present among the most strongly attracted collexemes. Even across the entire Spanish dataset, non-color adjectives only occur 9 and 5 times in the X and Y slot, respectively.

Turning to the covarying collexeme analyses, the top 10 attracted X-Y pairs are displayed in table 4. Again, there are striking similarities between the three languages, with the translation equivalents of *data is the new oil* occurring across all three languages. *Transparency is the new objectivity* and *small is the new big* are each attested in two of the three languages. The covarying collexemes are particularly informative about the ways in which the pattern is used to express perceived changes, especially social and cultural ones. Topics like digitization (*data are the new oil*, *blogs are the new resumé*), the extended period of youth and young adulthood (*preadolescents are the new adolescents*), and societal attitudes towards minorities (*muslims are the new jews*) transpire through the top 10 covarying collexemes.

A feature that sets apart Spanish from the other two languages is the prevalence of age terms (e. g., *30 is the new 20*, *40 is the new 30*) among the covarying collexemes, a fact that is also mirrored by the occurrence of *40* as one of the top-attracted simple collexemes both in the Spanish X and Y slot. Finally, an interesting observation is that despite the high frequency of *black* in the Y slot, with the exception of *gris es el nuevo negro* ‘gray is the new black’, no lexeme pair with *black* enters the top 10 attracted covarying collexemes. This speaks for a considerable degree of variability among the instances of [X *BE* the new *black*].

Table 5 shows the results of the cross-linguistic distinctive collexeme analysis. This approach uses multiple distinctive collexeme analyses to check which of the concept pairs formed by the lexemes in the X and Y slots are particularly attracted to each of the three languages. Given the differences in sample size across the three languages, and given that we use automatic translations of the German and Spanish lexemes, the results have to be taken with caution, but they do reveal interesting tendencies that confirm the results of our previous analyses. While all three languages show a fairly diverse set of concept pairs among their top 10 attracted collexemes (see the online supplementary material for more detailed results), Spanish shows a strong tendency towards age terms in the X and Y slots, along with terms from the domains of technology and culture. Terms referring to minorities in terms of, e. g., religion, sexual orientation, or ethnicity are most dominant in English

but also occur among the German top 10 terms. Color terms figure prominently among the English and German top 10, but not in Spanish. Overall, these results indicate that the use of the pattern is very similar in German and English and slightly more restricted in Spanish.

English			German			Spanish		
Lemmas	Freq.	T	Lemmas	Freq.	T	Lemmas	Freq.	T
<i>small-big</i>	75	1.29	<i>green-pink</i>	14	2.84	<i>30-20</i>	4	1.9
<i>blog-resume</i>	21	1.14	<i>uncool-cool</i>	11	2.66	<i>40-30</i>	4	1.9
<i>grey-black</i>	20	1.11	<i>blue-black</i>	17	2.49	<i>data-oil</i>	8	1.75
<i>muslims-jews</i>	15	0.97	<i>gray-black</i>	17	2.25	<i>transparency-objectivity</i>	5	1.68
<i>gay-black</i>	18	0.88	<i>blue-green</i>	10	2.16	<i>60-40</i>	3	1.64
<i>pink-black</i>	21	0.81	<i>reason-punk</i>	7	2.12	<i>content-advertising</i>	3	1.61
<i>comedy-rock'n'roll</i>	10	0.79	<i>data-oil</i>	21	2.04	<i>body-password</i>	2	1.34
<i>nra-kkk</i>	9	0.75	<i>autumn-summer</i>	6	1.97	<i>ecology-communism</i>	2	1.34
<i>big-small</i>	8	0.7	<i>lilac-black</i>	6	1.97	<i>fashion-rock</i>	2	1.34
<i>green-red</i>	22	0.69	<i>turk-jew</i>	6	1.97	<i>internet-mouth</i>	2	1.34

Table 5. Results of the cross-linguistic distinctive collexeme analysis; the German and Spanish lemmas were automatically translated into English prior to analysis

3.3 Discussion

Based on our corpus results, we can draw inferences about the status and productivity of [X BE the new Y] in each of our three target languages. Overall, the data support our hypothesis that the German and Spanish constructions are instances of “constructional borrowing” (Coleman, 2016) from English. First, the pattern is considerably less frequent in these two languages, suggesting that it has not achieved the same degree of conventionalization as in

English, and thus supporting the role of the latter as the source of an asymmetric transfer relationship.

Second, at least in Spanish, the variability of the slot fillers appears to be somewhat more constrained than in English: Apart from color terms, the Spanish construction does not typically combine with adjectives, and it seems to rely more heavily on recurrent semantic domains such as age (e. g., *30 is the new 20*). This lower degree of formal and semantic flexibility suggests that the Spanish pattern is used more conservatively than its English model.

Finally, specific (semi-)lexicalized subtypes are shared by all three languages: This includes fully lexicalized instances (e. g., *data is the new oil*) as well as the subpattern [X BE the new black], which may qualify as a subconstruction in its own right. The fact that these exact patterns are attested in each language suggests that they are the result of borrowing from a common source. In an alternative scenario in which [X BE the new Y] developed independently in each language, rather than through contact, such close lexical correspondences would not be expected (see Weber, 2019, for a similar argument).

Given the striking cross-linguistic parallels in the use of [X BE the new Y], a final question is how this multilingual relationship can be modelled from a constructionist perspective. Here, we build on Höder’s (2012, 2018) diasystematic construction grammar (introduced in section 2.2), in which language contact is assumed to lead to the formation of language-independent generalizations in the mental networks of multilingual speakers. Figure 1 illustrates a potential diasystem for the English and German snowclone; similar relations could be posited between the English and the Spanish construction.

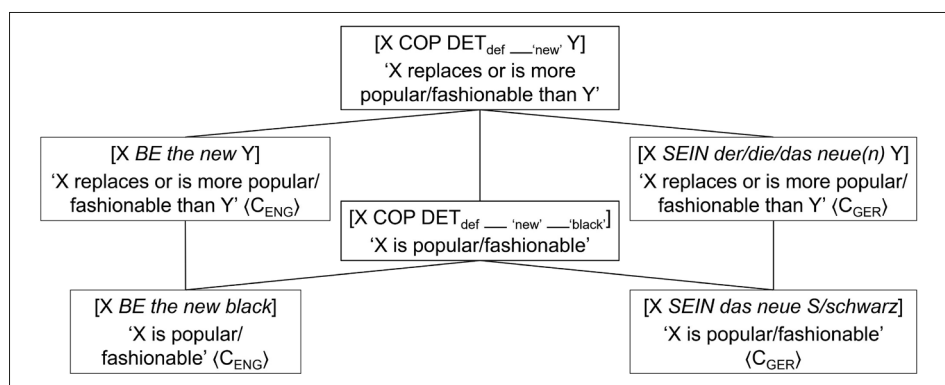


Figure 1. Diasystem of English and German constructions

The top center of figure 1 shows the cross-linguistic schema (or “diaconstruction”, in Höder’s terms) that speakers have inferred, which characterizes the formal elements and constructional meaning of the snowclone but leaves the lexical elements underspecified (e. g., COP is a placeholder for a copula verb, and ____{new} denotes a lexical item meaning ‘new’). Below this overall schema, the subpattern with an item meaning ‘black’ in the Y slot, which we have seen is productive across languages, is included as a lower-level diaconstruction. Rather than denoting a literal comparison with the color ‘black’, this subconstruction expresses the idiomatic meaning that X is generally popular/fashionable. Finally, the two sides of the diagram show the language-specific patterns (called “idioconstructions” in Höder’s framework), which contain lexically fixed elements and are subject to language-dependent communicative settings (denoted as $\langle C_{\text{ENG/GER}} \rangle$, following Höder’s [2018] notation). Crucially, the snowclones in each language, as well as their subtypes with *black/schwarz*, inherit their shared features from the diaconstructions, thus capturing cognitive efficiency in the multilingual system (Höder, 2012: 247-249).

It is an open question to what extent individual instances of the pattern may be direct translations from English. For instance, German *Daten sind das neue Öl* might be a direct translation of English *data is the new oil*, or the phrase could have been coined in both languages independently. To make things even more complicated, it may be a direct translation in some texts and a creative coinage by authors not familiar with the English phrase in others. This suggests that speakers’ cross-linguistic knowledge of the diaconstructional family is best modeled as a complex network at different levels of abstraction, ranging from specific constructs to subconstructions such as [X *BE the new black*] to the abstract pattern [X *BE the new Y*].

4. Conclusion

In this paper, we have presented a corpus-based analysis of the English construction [X *BE the new Y*] and its Spanish and German equivalents. The contrastive perspective can arguably contribute significantly to our understanding of “snowclones” as culturally salient phrasal patterns that can be reused even across multiple languages. In the case of [X *BE the new Y*], our analysis shows a number of cross-linguistic similarities that support the hypothesis that the pattern in its present-day usage emerged in English and spread from there to other languages. We have shown how this contact scenario can be

modeled via diasystematic links in multilingual speakers’ mental networks. Our findings illustrate that data from multi-language web corpora can provide a simple and efficient access route to tracing recent developments of phraseological patterns across languages.

Data availability

The datasets and the analysis script are available at <https://osf.io/t79c8/>.

References

- Andersen, Gisle. 2022. Phraseology in a cross-linguistic perspective: A diachronic and corpus-based account. *Corpus Linguistics and Linguistic Theory* 18(2): 365-389. doi: <https://doi.org/10.1515/cllt-2019-0057>
- Baayen, R. Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In Lüdeling, Anke & Kytö, Merja (eds.) *Corpus linguistics*. Berlin & New York: De Gruyter, 899-919.
- Boas, Hans C. & Höder, Steffen (eds.). 2018. *Constructions in contact: Constructional perspectives on contact phenomena in Germanic languages*. Amsterdam & Philadelphia: John Benjamins.
- Boas, Hans C. & Höder, Steffen (eds.). 2021. *Constructions in contact 2: Language change, multilingual practices, and additional language acquisition*. Amsterdam & Philadelphia: John Benjamins.
- Church, Kenneth W.; Gale, William; Hanks, Patrick & Hindle, Douglas. 1991. Using statistics in lexical analysis. In Zernik, Uri (ed.) *Lexical acquisition: Exploiting on-line resources to build a lexicon*. Hillsdale, NJ: Erlbaum, 115-164.
- Colleman, Timothy. 2016. A reflection on constructionalization and constructional borrowing, inspired by an emerging Dutch replica of the ‘time’-away construction. *Belgian Journal of Linguistics* 30(1): 91-113. doi: <https://doi.org/10.1075/bjl.30.05col>
- De Pascale, Stefano; Pijpops, Dirk; Van de Velde, Freek & Zenner, Eline. 2022. Re-assembling the pimped ride: A quantitative look at the integration of a borrowed expression. *Frontiers in Communication* 7. doi: <https://doi.org/10.3389/fcomm.2022.777312>
- Fiedler, Sabine. 2017. Phraseological borrowing from English into German: Cultural and pragmatic implications. *Journal of Pragmatics* 113: 89-102. doi: <https://doi.org/10.1016/j.pragma.2017.03.002>
- Flach, Susanne. 2021. *Collostructions: An R implementation for the family of collostructional methods*. <https://sfla.ch/collostructions> [Accessed 30/04/2024].
- Furiassi, Cristiano; Pulcini, Virginia & Rodríguez González, Félix (eds.). 2012. *The anglicization of European lexis*. Amsterdam & Philadelphia: John Benjamins.

- Gaeta, Livio & Ricca, Davide. 2006. Productivity in Italian word-formation: A variable-corpus approach. *Linguistics* 44(1): 57-89. doi: <https://doi.org/10.1515/LING.2006.003>
- Gilquin, Gaëtanelle. 2015. Contrastive collocation analysis: Causative constructions in English and French. *Zeitschrift für Anglistik und Amerikanistik* 63(3): 253-272. doi: <https://doi.org/10.1515/zaa-2015-0022>.
- Gries, Stefan Th. 2024. *Frequency, dispersion, association, and keyness: Revising and tupleizing corpus-linguistic measures*. Amsterdam & Philadelphia: John Benjamins.
- Hartmann, Stefan & Ungerer, Tobias. 2024. Attack of the snowclones: A corpus-based analysis of extravagant formulaic patterns. *Journal of Linguistics* 60(3): 599-634. doi: <https://doi.org/10.1017/S0022226723000117>
- Hartmann, Stefan. 2018. Derivational morphology in flux: A case study of word-formation change in German. *Cognitive Linguistics* 29(1): 77-119. doi: <https://doi.org/10.1515/cog-2016-0146>
- 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. doi: <https://doi.org/10.1080/00335630.2018.1519255>
- Höder, Steffen. 2012. Multilingual constructions: A diasystematic approach to common structures. In Braunmüller, Kurt & Gabriel, Christoph (eds.) *Multilingual individuals and multilingual societies*. Amsterdam & Philadelphia: John Benjamins, 241-258.
- Höder, Steffen. 2018. Grammar is community-specific: Background and basic concepts of Diasystematic Construction Grammar. In Boas, Hans C. & Höder, Steffen (eds.) *Constructions in contact: Constructional perspectives on contact phenomena in Germanic languages*. Amsterdam & Philadelphia: John Benjamins, 37-70.
- Jakubiček, Miloš; Kilgarriff, Adam; Kovář, Vojtěch; Rychlý, Pavel & Suchomel, Vít. 2013. The TenTen corpus family. In Hardie, Andrew & Love, Robbie (eds.) *7th International Corpus Linguistics Conference (CL2013)*. Lancaster: UCREL, 125-127.
- Onysko, Alexander. 2007. *Anglicisms in German: Borrowing, lexical productivity, and written codeswitching*. Berlin & New York: De Gruyter.
- Paquot, Magali & Bestgen, Yves. 2009. Distinctive words in academic writing: A comparison of three statistical tests for keyword extraction. In Jucker, Andreas; Schreier, Daniel & Hundt, Marianne (eds.) *Corpora: Pragmatics and discourse*. Amsterdam: Rodopi, 247-269.
- Pietsch, Lukas. 2010. What has changed in Hiberno-English: Constructions and their role in contact-induced change. *STUF - Language Typology and Universals* 63(02): 118-145. doi: <https://doi.org/10.1524/stuf.2010.0010>
- Pullum, Geoffrey K. 2004. *Snowclones: Lexicographical dating to the second*. [Blog post]. <http://itre.cis.upenn.edu/~myl/languageelog/archives/000350.html> [Accessed 30/04/2024].

- R Core Team. 2024. *R: A language and environment for statistical computing*. [Manual]. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/> [Accessed 30/04/2024].
- Rühle, Alex. 2010. *Erschütternd ist das neue geil*. [Newspaper article]. <https://www.sueddeutsche.de/kultur/trend-wendung-im-deutschen-erschuetternd-ist-das-neue-geil-1.420637> [Accessed 30/04/2024].
- Schäfer, Roland. 2015. Processing and querying large web corpora with the COWI4 architecture. In Bański, Piotr; Biber, Hanno; Breiteneder, Evelyn; Kupietz, Marc; Lungen, Harald & Witt, Andreas (eds.) *Proceedings of the 3rd Workshop on Challenges in the Management of Large Corpora (CMC-3)*. Mannheim: Institut für Deutsche Sprache, 28-34.
- Schäfer, Roland & Bildhauer, Felix. 2012. Building large corpora from the web using a new efficient tool chain. In Calzolari, Nicoletta; Choukri, Khalid; Declerck, Thierry; Uğur Doğan, Mehmet; Maegaard, Bente; Mariani, Joseph; Moreno, Asuncion; Odijk, Jan & Piperidis, Stelios (eds.) *Proceedings of the Eighth International Conference on Language Resources and Evaluation*. Istanbul: European Language Resources Association, 486-493.
- Stefanowitsch, Anatol & Gries, Stefan Th. 2003. Collocations: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2): 209-243. doi: <https://doi.org/10.1075/ijcl.8.2.03ste>
- Stefanowitsch, Anatol & Gries, Stefan Th. 2005. Covarying collexemes. *Corpus Linguistics and Linguistic Theory* 1(1): 1-43. doi: <https://doi.org/10.1515/clt.2005.1.1.1>
- Traugott, Elizabeth Closs & Trousdale, Graeme. 2013. *Constructionalization and constructional changes*. Oxford: Oxford University Press.
- Van de Velde, Freek & Zenner, Eline. 2010. Pimp my Lexis: het nut van corpusonderzoek in normatief taaladvies. In Hendrickx, Els; Hendrickx, Karl; Martin, Willy; Smessaert, Hans; Van Belle, William & Van der Horst, Joop (eds.) *Liever meer of juist minder? Over normen en variatie in taal*. Gent: Academia Press, 51-68.
- Wasserscheidt, Philipp. 2014. Constructions do not cross languages: On cross-linguistic generalizations of constructions. *Constructions and Frames* 6(2): 305-337. doi: <https://doi.org/10.1075/cf.6.2.07was>
- 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. doi: <https://doi.org/10.13092/lo.96.5535>
- Zenner, Eline; Heylen, Kris & Van de Velde, Freek. 2018. Most borrowable construction ever! A large-scale approach to contact-induced pragmatic change. *Journal of Pragmatics* 133: 134-149. doi: <https://doi.org/10.1016/j.pragma.2018.05.004>

