

Damien Villers

Université de Toulouse-Jean Jaurès

France

<https://orcid.org/0000-0002-6382-4731>

Phraseological Blunders: When New Phrasemes Are Born from Errors

Abstract. This article first presents a typology of phraseological “blunders” – which include both errors and mistakes that accidentally modify the standard form, usage, or meaning of a phraseme – and then outlines the causes and the types of interference behind them, such as language pathologies, poor linguistic skills, or mere absence of mind. It then proceeds to study the notion of variation in order to draw a line with the notion of error with the help of criteria such as frequency ratio, communicational efficiency, and semantic coherence. Finally, the article presents cases of phraseological accidents that have lexicalised due to various cognitive biases, thus becoming new phrases. In order to account for this seemingly paradoxical phenomenon, the memetic approach is selected to build a presentation of the “phraseme genesis” process (or “phraseologisation”) and the selection criteria that facilitate it.

Key words: *phraseme, error, variation, propagation, memetics*

1. Introduction

As a result of their prefabricated nature, phraseological units (PUs) exhibit some formal stability, and are thus prone to various modifications. It is important to stress from the very start that this study will not deal with deliberate or premeditated modifications. Such modifications have been studied extensively and are generally called “creative idiom modifications”, “anti-idioms”, “phraseological puns”, “idiom parodies”, “twisted phrasemes” and so forth. This study will only tackle phraseological modifications that are the result of unintentional deviation and that do not match the attested form, usage, or meaning of a phraseme. As for the term “phraseme”, it will be used

interchangeably with “phraseological unit”¹ to refer to all² preconstructed³ polylexical units. This definition implies that the traditional view was chosen in this study. Let’s consider the following inappropriate utterances:

- (1) I tell people, let’s don’t fear the future, let’s shape it!* (G. W. Bush, Omaha, 2006).
- (2) They misunderestimated me* (G. W. Bush, Bentonville, 2000).
- (3) Being from a family of outlaws makes you a social leopard* (*The Law of Finders Keepers*, 2018).

Some scholars have recently argued in favour of a wider conception of phraseology that includes even polylexemic units, i.e. patterns and compounds. Under that view, the mistakes in examples (1) and (2) should be incorporated in the present study. The former is a blend of *let’s not* and *don’t* while the latter seems to be a mix of *misunderstand* and *underestimate*. However, under the traditional view, only example (3), where *leper* was replaced with its near homonym *leopard*, will be considered a phraseological blunder. As for examples (1) and (2), they will be considered syntactic and lexical blends.

Bergstrom (1906), Bolinger (1961), Cohen (1987), and Legallois (2013) are among the very few who published articles and volumes on syntactic blends – and also lexical blends in the case of Bergstrom. Works on “phraseological errors” are more recent and appeared in the 2000s due to the boom of phraseodidactics, the study of phraseological competence for language teaching and learning. Among notable works, it is possible to cite Nesselhauf (2003), Osborne (2008), Paquot (2008), Thiessen (2008), or Wang & Shaw (2008). They all focus on erroneous collocations among L2 English learners and the influence of their native language (L1) on such errors. From a less didactic and more linguistic viewpoint, Polguère (2007) studied the nature of “collocational grafts” (or “blends”), a type of phraseological error; while Liudmila Liashchova gave a presentation at Europhras 2018 about the “phraseological errors” made in Russian by a highly fluent US journalist.

¹ Other famous general labels include “formulaic language”, “fixed expressions”, “multi-word units”, “set-phrases”, “idioms” (which may also refer to a non-compositional subclass of phraseme), or “lexical bundles” among computational linguists.

² A minority of scholars, especially in Russia and Slavic countries, use the term to refer to figurative set phrases only. This is due to the influence of A.V. Kunin’s works and terminology, which were greatly influenced by Charles Bally’s *Traité de Stylistique Française* and its terminology.

³ From a cognitive viewpoint, other common terms include “reproducible”, “formulaic”, and “prefabricated”. From a structural point of view, they correspond to the notions of “frozenness” and “fixedness”, or the less aggressive notion of “stability”.

She analysed from a linguistic-pragmatic viewpoint some of his mistakes and the reasons behind them. Finally, it is worth noting that a website dedicated to idiom-based blends – “malaphors⁴” – appeared in 2012 (malaphors.com). It serves as a database and a forum where users list and discuss idiom blends gathered in the media.

2. Typology of phraseological blunders

Several typologies of phraseological errors have been formulated by scholars. According to James (1998), phraseological errors may be classified into two general categories: “grammaticality errors” (incorrect form) and “acceptability errors” (inappropriate usage). Thiessen (2008: 5–6) proposes a classification that is based on (some) phraseme types: lexical collocation errors, grammatical collocations errors, errors in idiom-like phrases, and phrasal verb errors. Finally, Liashchova (2018) notes that the major deviations are excessive use and inadequate use. She subdivides the latter into erroneous synonymy, loan translation or calque, comprehensible blending and incomplete semantization. While very useful, these typologies are either very general or focus on very specific approaches (phraseme type or semantics) and do not account for all existing types of phraseological accidents. Therefore, the following classification is proposed:

- Substitution

(3) Being from a family of outlaws makes you a social leopard* (*The Law of Finders Keepers*, 2018).

- Permutation

(4) Look who’s calling the pot black!* (*NCIS*, 3–18).

- Expansion

(5) I’m so fed up with being the escape goat for all the problems of my bloody family* (wehavekids.com).

- Omission

(6) I personally could care less if the shirt is made in the U.S. or not. That doesn’t matter to me* (www.mintees.com).

⁴ The term itself is a blend of *malapropism* and *metaphor*.

- Fusion or “blending”

(7) The whole Transformers thing isn’t my kettle of tea* (eurobricks.com).

- Inter-language calque

(8) You’re selling the bear’s skin before you’ve killed it!*

- Inappropriate contextual use.

(9) I know my lesson off the top of my head*.

- Wrong interpretation (absence of comprehension, semantic blends, literal reading etc.)

- Overuse and underuse.

- Mixed type

(10) Families is where our nation finds hope, where wings take dream* (G. W. Bush).

Substitution, exemplified by (3), is probably the most common type of phraseological blunder. In this example, the idiom *to be a social leper* (to be shunned) is modified by substitution of the term *leper* with its near homonym *leopard*. Such Homonymy-based errors are often called “eggcorns⁵”. Example (4) is a modification of *to call the kettle black*, which is a variant of *the pot calling the kettle black* and refers to hypocritical criticism. In this case, the terms *pot* and *kettle* are substituted, reversing their order. Such permutations are a rare type of malaphor and could be seen as a subtype of substitution. In addition, it must be noted that instances (3) and (4) are errors from the character’s viewpoint only since they are intentional for the writer. In (5), the word *scapegoat* is turned into a collocation that does not exist, thus increasing its lexical volume. Naturally, the expansion category does not only contain words that are turned into phrasemes, but may also concern phrasemes that are made even longer by adding letters or words to them. It is the case with *to exact revenge* and *to set foot on*, which are often misused as *to extract revenge** and *to step foot on**. Example (6) is the exact opposite, as the conversational routine *I couldn’t care less* was made shorter by omitting the term *not* or the phonemes /n(t)/. Another common mistake is *for all intensive purposes**, which is a lexical reduction of *for all intents and purposes*. Blending⁶ or “fusion”, exemplified by (7), is a common type of error, at least the most studied

⁵ The term is generally attributed to linguist Geoffrey Pullum, who reportedly suggested the label after reading an article about a patient who substitutes the term *eggcorn* for *acorn*.

⁶ Aarts (2007: 189), who studies syntactic blends, further distinguishes between the more integrated “blends” and “mergers”, where two distinct components can still be identified.

in literature. In this instance, the idioms *kettle of fish* (a tricky situation) and *not my cup of tea* (to dislike something) were blended together. Utterance (8) is a case of inter-language calque, where a phrase is translated word by word into a language in which it does not exist. Here, the phrase in question is *vendre la peau de l'ours avant de l'avoir tué*, whose correct English equivalent is *to count your chickens before they've hatched*.

Unsurprisingly, PUs – mostly non-compositional ones – may also be misunderstood or misused. Such “acceptability errors” (James 1998) are extremely frequent, especially among learners, which explains why the vast majority of studies on phraseological errors are language-learning oriented. Inappropriate usage is exemplified by (9), where the idiom *off the top of my head* (without thinking) is used in a wrong context, where *by heart* would have been more suitable. The next category, wrong interpretation, is probably the most common type, especially in translation exams. For instance, second-year students at the university of Toulouse were given a news article to translate into French for their exam. It contained an excerpt about a big company's performance on the Dow Jones industrial average, in which the said company was referred to as a “blue-chip conglomerate”. Out of 23 students, 21 failed to get the correct meaning in their translations, opting instead for omission, loan translations, nonsensical phrases, false meanings, barbarisms, and so forth. This shows that wrong interpretation could be further divided into several types that include the absence of comprehension as well as erroneous comprehension, which, in turn, comprises semantic blends with another phraseme or literal reading. Another type of blunder is overuse and underuse. Paquot (2008) gives the example of *for instance* and *for example*, which are overused among L2 English learners when expressing exemplification. Finally, the typology would be incomplete without a category for hybrids. Some blunders, as in (10), are a mix of several types. Let's suppose for a moment that George W. Bush meant “where dreams grow wings”. This would imply that he mixed up *to grow wings*, *to take flight* and reversed the word order. Similarly, *to step foot on** may be viewed as an addition (of one letter), or as a blend of *set foot on* and *step on*, or even as a substitution.

3. The causes behind phraseological blunders

As stated before, the majority of studies on phraseological errors deal with language teaching and learning. Therefore, they focus on errors made by learners and non-natives (e.g. Nesselhauf 2003, Osborne 2008, Paquot 2008, Thiessen 2008, or Wang & Shaw 2008). Unsurprisingly, these studies show

that their native tongues have a lot of influence on errors in L2 English and that insufficient knowledge of a language is the main cause for phraseological errors among learners. However, other causes may be put forward: as Granger (2004: 135) puts it, “advanced interlanguage is the result of a very complex interplay of factors: developmental, teaching-induced, and transfer-related”. In her presentation on errors made in Russian by a non-native US journalist, Liashchova (2018) also noted that incorrect use is due to insufficient learning. As for phraseme overuse, she posits that it is the result of personal affection for idioms, a desire to make use of their pragmatic functions, as well as a desire to sound authentic. In his PhD thesis on “frozenness”, Misri (1987: 414) claims that accidental modifications are common “among children who have not yet mastered the linguistic system, foreigners who have limited knowledge of frozen units, and patients with language-related pathologies”.

Many studies⁷ confirm that proverbs and idioms are used in various tests to detect pathologies such as dementia, Schizophrenia, Right Hemisphere Damage, Alzheimer’s, or aphasia. These tests include IQ tests (*Wechsler Adult Intelligence Scale / WAIS-R, Stanford-Binet test*), personality tests (*Famous Sayings Test, Attitude Measurement Test*), executive functions tests (*Delis-Kaplan Executive Function System, Proverb Interpretation Task*), or psychopathology tests (*Gorham Proverb Test*). The omnipresence of proverbs – at least figurative⁸ ones – and idioms in such tests is due to their neurocognitive complexity. As explained by Honeck (1997: 220–222) and his DARTS model, both brain hemispheres are required to process them, which makes them very valuable from a medical point of view.

Nonetheless, reducing phraseological blunders to pathologies or insufficient learning is excessive, as they only account for a fraction of such blunders. As a matter of fact, the vast majority of native speakers, including experienced and healthy speakers, are prone to such errors. According to Bergstrom (1906), who studied syntactic and lexical blends, blunders are caused by some sort of “contamination”; while Polguère (2007), who studied collocational blends (or “grafts”), claims they are caused by an “interference”. This explanation may be applied to all types of phraseological blunders but these phenomena should be seen as intermediate causes, and not the root cause of the problem. Both scholars agree that contamination or interference

⁷ Most of these studies on neurological applications for proverbs and idioms are summarised in Van Lancker (1990) and Murphy et al. (2013).

⁸ Paremiologists are divided as to whether metaphor should be considered optional or obligatory in proverb definition. This point is debated in Villers (2014), along with other criteria.

can in turn be explained by the notion of *analogy*. In other words, phraseological blunders are due to a similarity with a sound, a lexical element, a theme, or a meaning found in another phraseme – including phrasemes from other languages. Once again, these phenomena may be seen not as the root cause of the problem, but as intermediate causes. Furthermore, analogy is a very common⁹ cognitive process, at the heart of language and idiom processing and learning. Since analogy is so omnipresent, more specific triggers need to be put forward. Such triggers have actually long been identified in cognitive and psycholinguistic studies¹⁰ on performance errors and are generally labelled “performance factors” after Chomsky’s *Aspects of the Theory of Syntax* (1965). They include lack of attention, memory lapses, tiredness, lack of interest, emotional state, drugs and alcohol, etc. Insofar as these factors apply to all types of speech, they are, of course, valid for phraseological blunders as well. In didactics, the label “error” generally implies systematic deviation owing to competence factors while “mistakes” entail temporary performance factors. Therefore, the term “blunder” will be used to encompass both “mistakes” and “errors”.

4. The blurred line between error and variation

Dealing with the notion of error instantly raises the difficulty of setting boundaries with the notion of variation. Phraseological blunders were previously described as utterances that are the result of unintentional deviation and that do not match the attested form, usage, or meaning of a phraseme. Yet, some deviations are hard to classify, especially the ones that occur frequently. In this case, two ideologies clash: descriptivists will consider that common or systematic deviations should not be viewed as errors but as idiosyncrasms, while prescriptivists will view them as erroneous. It is therefore necessary to find a more objective and intermediate stance that incorporates all relevant criteria in order to establish functional delimitations. In corpus phraseology, the minimum frequency of (co)occurrence for a string of words to be considered preconstructed or “reproducible” – and thus deserve the label

⁹ The same could be said of blends. According to the “conceptual blending” theory, designed by Fauconnier and Turner (2002), blends are an omnipresent mental process; they are at the centre of how we create meaning and how we think. This theory is similar to George Lakoff’s “conceptual metaphor” theory, according to which we see the world through metaphors.

¹⁰ Among major studies on performance errors, one may cite Crain & Thornton (1998, chap. 15), Gleason & Ratner (1993), or Kamhi (1988).

of “phraseme” – is generally described as being higher “than would be expected by chance”. This distinction could not be applied to errors related to preconstructed language units, as both erroneous and standard forms are bound to be recurrent. When it comes to errors derived from free combinations, there is no known quantitative threshold either. This means that phraseological blunders cannot be singled out by means of frequency alone. Other criteria must be taken into account, as will be revealed through the examples in the table.

Table 1. Examples of variants and errors with their frequency in online corpora

Utterances and possible variants/errors	Google Books corpus (189 bn words)	iWeb ¹¹ corpus (14 bn words)
(3a) A social leper	5,770/610	32
(3b) A social leopard*	151	0
(4a) (Pot) Calling the kettle black	49,600	695
(4b) (Pot) Calling the pot black*	2,560	25
(5a) A scapegoat	3,440,000	9415
(5b) An escape goat*	2,600	88
(6a) I couldn't care less	58,620	1,111
(6b) I could care less*?	66,700	3,526
(11a) The early bird gets the worm	8,170	367
(11b) The early bird catches the worm	13,100	257
(12a) Nothing ventured, nothing gained	44,900	118
(12b) Nothing ventured, nothing had*	132	0
(13a) For all intents and purposes	561,000	7,663
(13b) For all intensive purposes*	2,990	381
(14a) To exact revenge	71,900	1,600
(14b) to extract revenge*	4,420	92
(15a) To set foot (on)	88,000	14,190
(15b) to step foot (on)*	4,510	3,973
(16a) First come, first served	858,000	3,535
(16b) First come, first serve*?	70,800	3,847
(17a) A leopard cannot ¹² change its spots	7,466	19
(17b) A tiger cannot change its stripes	1,099	2
(17c) A zebra cannot change its stripes	257	1
(17d) A zebra cannot change its spots*?	494	2

Source: own research.

¹¹ Can be found on corpus.byu.edu along with other corpora compiled by Mark Davies.

¹² All full and contracted forms were included: *cannot*, *can't*, *does not*, *doesn't*.

The first three examples confirm that the frequency criterion alone is not reliable since even the most ludicrous mistakes are significantly frequent. Moreover, frequency numbers need to be taken with a pinch of salt insofar as some utterances may correspond to quotations or titles, not to mention the margin of error inherent to Google Books. This is precisely why (10), “wings take dream”, now a famous Bushism¹³, cannot be studied, as it was quoted and commented on at length. It even inspired book titles. In fact, the *ratio* between the standard form and the deviation should be considered the most relevant frequency-related criterion.

In the case of (5), where the ratio is 1320 to 1 and 106 to 1 in favour of (5a), this criterion is enough to declare (5b) erroneous. The ratio is much higher in the case of (3), with a proportion of 38 to 1, but it is still low enough to make (3b) a phraseological blunder, unless uttered in a documentary on leopards. As a matter of fact, the ratio for both (3b) and (5b) is so low compared to (3a) and (5a) that their frequency curves are not visible on the n-gram viewer, the analytical tool based on Google Books. From a semantic viewpoint, it is interesting to note that even if these errors are very amusing, they are not entirely illogical. Not only are they close homonyms, they also contain a certain degree of semantic motivation: leopards do tend to be avoided, and goats might allow one to escape. Besides, *scapegoat* was diachronically derived from *escape goat*. Example (4) is different: although its low ratio of 19 to 1 and 28 to 1 (cf. Figure 1) is sufficient to claim that (4b) is erroneous, another filter may be added: pun probability. As it turns out, a good proportion of contexts involving (4b) indicate that the deviation – which implies the repetition of *pot* – is deliberate, either for emphasis or humorous intent. When taking this factor into account, the ratio appears to be even lower for genuinely deviant uses. The same may be said of (3b) and (5b), which is why the figures obtained from corpora need to be put into perspective. To conclude, (4b) should be considered erroneous if it accidentally deviates from the idiom in (4a).

Examples (11) and (12) are probably the easiest cases. Both proverb variants in (11) are correct and attested in numerous dictionaries and occur quite proportionally in both corpora. By contrast, (12b) is not an attested variant of the proverb in (12a) due to its very low ratio (340 to 1). Deliberate modification is very unlikely in the case of (12b), which might account for such low figures. The next two examples remain fairly easy to classify but

¹³ This term commonly refers to a malapropism uttered by Georges W. Bush. His linguistic faux-pas were so common that a name was soon coined for them. There are even websites dedicated to them, and they even appear on tee shirts, mugs, posters, and so on.

Figure 1. Frequency curves for (4a) and (4b) on Google Books' n-gram viewer



Source: own research.

the frequency ratio is no longer the only relevant criterion. (13b) and (14b) both show a very low ratio¹⁴: 187 to 1 and 20 to 1 for (13), while (14b) is proportionally 16 and 17 times less frequent than (14a). From a semantic viewpoint, one may argue that the terms *intensive* and *extract* are nonsensical and illogical, even figuratively: how can a purpose be intensive; and why would revenge need to be extracted from something? Furthermore, these variants most probably originated in accidental phonological deformation or slips, as is often the case in idioms. All these criteria allow us to conclude that (13b) and (14b) are erroneous and should not be viewed as acceptable collocations, unless the deviation from standard form is intended as humour, sarcasm, a pun, or a quotation in order to comment on it.

Example (15b) is more difficult to analyse. First of all, its ratio varies greatly from one corpus to another. Thus, corpus type and speech quality should be taken into account. On the one hand, it shows a very low ratio (20 to 1) on a gigantic corpus of published books such as Google Books. On the other, it shows a much higher ratio (3.5 to 1) on a smaller corpus composed of websites and forums (iWeb), which are likelier to contain errors and loose language. It is also important to remember that some of the utterances from the corpora are meant to be sarcastic comments, which further balances this seemingly good ratio. In addition, the frequency curve obtained from the overall percentage for (15b) in the whole Google Books corpus is nearly inexistent on the n-gram¹⁵ viewer, compared to the curve for (15a) (cf. Figure 2).

¹⁴ With 1 representing [13b]. The potentially erroneous variant is always represented by 1 in the present article.

¹⁵ The n-gram viewer may show slightly skewed results due to its margin of error, as it is based on Google Books.

Figure 2. Frequency curves for *set foot* and *step foot* on Google Books' n-gram viewer



Source: own research.

Another argument is the semantic redundancy contained in *step foot*, given that the verb *step* already implies moving one's *foot* or *feet* and placing it or them on a new surface. Regarding its origins, the phrase is most certainly the result of an accidental graft or "blend" – between the collocation *to set foot* and the verb *to step (on)* – that was made possible due to phonological and semantic resemblance. For all these reasons, we argue that (15b) should be considered erroneous when used without satirical or humorous intent.

(16) poses a similar challenge as its ratio varies greatly from one corpus to another: 12 to 1 for Google Books and 1 to 1.1 for iWeb. Even when taking into account corpus type or speech quality and the rather low ratio of 12 to 1, the evidence against (16b) is scarce. A contextual analysis reveals that 58% of occurrences for (16b) include cases where the proverb is used as an adjectival phrase: *on a first-come first-served basis*. This suggests that one of the main factors that led to its emergence is phonological economy, in order to avoid specific phonemic combinations (viz. /d/ and /b/ without pause). The semantic criterion is of great help in this case as it reveals that (16b) is semantically incoherent since it states something different from what is meant (serving other people instead of being served). One may object that the active voice may be coherent from the perspective of a salesperson or a merchant ("serve those who have come first"), to which could be replied that the perspective of the beneficiary is the more logical one for the latter is the logical subject. Ultimately, this lack of semantic clarity violates the Gricean maxim of manner and may create misunderstandings or ambiguity, which is why it should be considered erroneous. For the same reasons, (17d) should not be considered a valid variant of (17a) given that zebras, as a rule, have no spots. Besides, the fact that it has more hits in corpora than

Figure 3. Frequency curves for (6a) and (6b) on Google Books' n-gram viewer



Source: own research.

a bona fide variant such as (17c) is attributable to a famous mistake, made by Al Gore, which was heavily commented on.

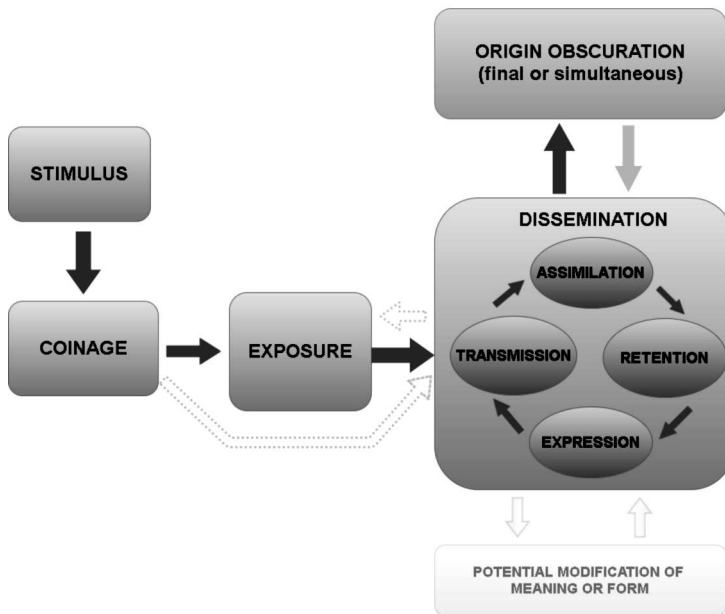
The most challenging case was saved for the end. In (6), the deviant and phonetically reduced form *I could care less* is more frequent than the standard one, or seems¹⁶ to be. While its ratio works in favour of (6b) in both corpora, the n-gram viewer – based on Google Books – shows contradictory results (cf. Figure 3). Moreover, a quick contextual analysis reveals that a large number of uses of (6b) turn out to be extracts where its grammaticality is debated. But even when balancing this ratio by filtering out grammatical comments and only keeping actual uses, *I could care less* is still very frequent, especially in the United States. The main criterion usually put forward to deny its grammaticality is the semantic one: the phrase actually says the very opposite of what it actually means. Unlike the previous examples from the table, (6b) is defended by several scholars, who make little of this semantic argument owing to the non-compositional nature of idioms. Among them is Stephen Pinker, who claims in *The Language Instinct* (p. 377) that *I could care less* is sarcastic, or John Lawler and Mark Lieberman who claim on their blogs that the phrase is, rather, a case of “negation by association”, where a construction retains its negative force without negation markers. While prosody might sustain these arguments – in some cases only, the notion of communicative efficiency is more important. In line with this, (6b) should be considered erroneous as it may create ambiguity and misunderstandings when used figuratively to mean the opposite of what its literal level or “surface structure” reads.

¹⁶ Naturally, both contracted and full forms were taken into account.

5. Phraseme genesis and selection criteria

The fact that illogical or erroneous variants can catch on and lexicalise no longer needs proving. It now needs explaining. The answer to this paradoxical phenomenon lies in the phraseme genesis process, which was described at length in Villers (2015, 2016, 2018).

Figure 4. The stages of phraseme genesis, adapted from proverb genesis



Source: own research.

In a nutshell, a stimulus (a specific situation) triggers the coinage of a phrase which in turn needs to be exposed to a sufficient number of hosts (speakers) – or vice versa – with the help of a propagating agent (a film, a website, a book, etc.). By repeating the phrase, the speakers perpetuate the cycle and allow the phrase to propagate in a virus-like pattern. This cycle may be broken into separate stages, inspired by Heylighen (1998) and its memetic¹⁷ approach: in order to spread, the phrase first needs to be noticed, understood, and accepted (assimilation). It then has to be remembered and

¹⁷ Memetics is the study of how cultural units (or “memes”, such as trends or customs) replicate by imitation. This discipline, which is the cultural equivalent of genetics, was heavily criticised for its lack of results and was soon buried, although breakthroughs came later (neurological evidence and concrete applications).

used through a medium (i.e. voice, ink, a signal, etc.). The chances of replication – and therefore of survival¹⁸ – are very low for most phraseme candidates, as the majority of potential hosts (speakers) do not repeat the cycle. In fact, the odds greatly depend on “selection criteria”. Heylighen (1998) describes these selection criteria for the replication of memes and cultural units in general, which were modified and adapted to phrasemes in Villers (2018). Thus, the main selection criteria in phraseme genesis and replication are:

- Novelty: a phraseme is likelier to replicate if it is perceived as being new, especially among younger age groups.
- Originality: the odds for replication are better if it has no direct “competition”.
- Stylistic markers: a phraseme that “stands out” from free combinations by signalling its preconstructed nature has more chances to be noticed, used, and remembered.
- Simplicity: a PU will be probably be used more often and will spread faster if it is easily pronounced or understood.
- Usefulness: a phraseme is likelier to catch on if it is applicable to many situations.
- Authority: a PU associated with a famous person or work will spread faster, even if that source is progressively forgotten.
- Conformity: a PU needs to be accepted in order to be repeated, which implies that it cannot go against the beliefs or knowledge of hosts.
- Publicity: most importantly, the phraseme candidate needs mass exposure in order to reach an optimal number of hosts (speakers). Publicity greatly influences its scope of dissemination (local, national, etc.).

Naturally, some criteria have more weight than others. In theory, the “conformity” criterion should prevent or limit the propagation of erroneous variants. In fact, it is in competition with other criteria. Some phraseological blunders might, for instance, be deemed more appealing due to their apparent novelty, at least among some speaker groups. Some might also be deemed simpler, as phonological reductions can make them easier to pronounce (e.g. 16b). Publicity may, of course, help a blunder “catch on”, since a mistake with a lot of exposure is likelier to be repeated, as with (17d). Even the “conformity” criterion might cause a host to favour an erroneous variant over a standard one, as with (5b), which might seem more logical to speakers who have never heard the word *scapegoat*. In other words, the whole process and its

¹⁸ The replication process is very selective and has a low-success rate, in a “survival-of-the-fittest” manner.

selection criteria leave room for error insofar as the “hosts” or “vehicles¹⁹” responsible for the replication of phrasemes are human speakers, who are prone to cognitive biases and may therefore not detect the deviation during the assimilation phase, or deem it acceptable.

6. Conclusion

What transpires from studies on phraseological errors is that scholars focus on specific categories of phrasemes (collocations) and certain categories of errors (blends and substitutions). However, the typology that was presented revealed that phrasemes may be accidentally modified in numerous ways, be it their form, their usage, or their meaning. Phraseological blunders are attributable to several core reasons: absence of mind and emotional state are the most common triggers for mistakes among experienced speakers and natives, while the lack or loss of linguistic competence is the most common reason behind phraseological errors among learners and people with language-related pathologies. The study of several recurrent phraseological blunders highlighted the difficulty of drawing a line between error and variation; a set of criteria was hence used to assess and classify them: frequency ratio between standard form and deviation, context (to filter out quotations and deliberate modifications), semantic coherence, and communicative efficiency.

The examples under study also demonstrated the possibility for “erroneous” variants to lexicalise, leading to what may be seen as a case of corrupt innovation. This seemingly paradoxical phenomenon was explained with a memetic approach of the process of phraseme genesis and propagation, during which cognitive biases are not always filtered out by the various selection criteria. Indeed, grammaticality, logic, or coherence are sometimes non-essential factors to human hosts – to deviate is human. This phenomenon may be compared to the propagation of fake news or rumours, where coherence and truth and not necessarily deemed important by the hosts who propagate them. Ultimately, the existence of phraseological blunders, some of which are recurrent, was to be expected – not only from a statistical perspective, but also owing to human nature.

¹⁹ This term is the one used in memetics, while “host” is the used in epidemiology. It corresponds to speakers.

Bibliography

- Aarts, Bas. 2007. *Syntactic gradience*, Oxford university press, Oxford.
- Bergstrom, Gustaf .1906. *On blendings of synonymous or cognate expressions in English*. Lund: Hakan Ohlsson
- Bolinger Dwight. 1961. Syntactic blends and other matters. *Language* 37, 366–381.
- Crain, Stephen & Thornton, Rosalind. 1998. *Investigations in Universal Grammar*, Cambridge: MIT.
- Cohen, Gerald. 1987. *Syntactic Blends in English parole*. Verlag Peter Lang.
- Fauconnier, Gilles & Turner Mark. 2002. *The Way We Think. Conceptual Blending and the Mind's Hidden Complexities*, Basic Books, New York.
- Gleason, Jean Berko & Ratner, Nan. 1993. *Psycholinguistics*, New York: Harcourt Brace Jovanich.
- Granger, Sylviane. 2004. Computer learner corpus research: current status and future prospects. In Connor, Ulla; Upton, Thomas. (eds.). *Applied Corpus Linguistics: A Multidimensional Perspective*. Amsterdam & Atlanta: Rodopi. 123–145.
- Heylighen, Francis. 1998. What makes a meme successful? Selection criteria for cultural evolution. *Proc. 15th Int. Congress on Cybernetics*, 418–423.
- Honeck, Richard. 1997. *A Proverb in Mind*, LEA, London.
- James, Carl. 1998. *Errors in Language Learning and Use*. London and New York: Longman.
- Kamhi, Alan G. 1988. A Reconceptualization of Generalization and Generalization Problems. *Language, Speech, and Hearing Services in Schools* v19, 304–13.
- Liashchova, Liudmila. 2018. Phraseological errors in a foreigner's talk discourse and their reasons. Europhras conference 2018 (Bialystok).
- Legallois, Dominique. 2013. Les greffes phraséologiques – ou quand la syntaxe se compromet. *Langages* n° 189, 103–120.
- Misri, Georges. 1987. *Le figement linguistique en français contemporain*, PhD diss., Université Paris V.
- Murphy, Patrick *et al.* 2013. Impairments in proverb interpretation following focal frontal lobe lesions, *Neuropsychologia*, n° 51, 2075–2086.
- Nesselhauf, Nadja. 2003. The Use of Collocations by Advanced Learners of English and Some Implications for Teaching. *Applied Linguistics* 24 (2), 223–242.
- Osborne, John. 2008. Phraseology effects as a trigger for errors in L2 English: The case of more advanced learners. In Meunier, Fanny; Granger, Sylviane. (eds). *Phraseology in Foreign Language Learning and Teaching*, Amsterdam/Philadelphia: John Benjamins, 67–83.
- Paquot, Magali. 2008. Exemplification in learner writing: A cross-linguistic perspective, in Granger, Sylviane; Meunier, Fanny (eds). *Phraseology: An Interdisciplinary Perspective*. Amsterdam and Philadelphia: John Benjamins.
- Polguère, Alain. 2007. *Soleil insoutenable et chaleur de plomb : le statut linguistique des greffes collocationnelles*, *Academia* (revised official version).

- Thiessen, Jenifer. 2008. The phraseological errors of French, German, and Spanish-speaking EFL learners: Evidence from an error-tagged learner corpus, In Associação de Estudos e de Investigação Científica do Isla-Lisboa (eds). *Proceedings from the 8th Teaching and Language Corpora Conference (TaLC8)*, 300–306.
- Van Lancker, Diana. 1990. The Neurology of Proverbs, *Behavioural Neurology*, 169–187.
- Villers, Damien. 2014. *Le Proverbe et les Genres Connexes*, Presses Académiques Francophones, Sarrebruck.
- Villers, Damien. 2015. Proverbiogenèse et obsolescence, la naissance et la mort des proverbes, *Proverbium : Yearbook of International Proverb Scholarship*, n° 32, 383–424.
- Villers, Damien. 2016. Proverbiogenesis: the phases of proverbial birth, *9th Interdisciplinary Colloquium on Proverbs – Proceedings, Tavira: International Association of Paremiology*, 369–380.
- Villers, Damien. 2018. *The cake is a lie: witnessing the birth of a modern phraseme*, In Filatkina, Natalia & Stumpf, Sören (eds). *Konventionalisierung und Variation*, Peter Lang: Berlin.
- Wang, Ying & Shaw, Philip. 2008. Transfer and universality: Collocation use in advanced Chinese and Swedish learner English. *ICAME Journal* 32, 201–228.

Gaffes phraséologiques: quand de nouveaux phrasèmes naissent d’erreurs

Résumé

Le présent article propose en premier lieu une typologie des erreurs phraséologiques, où le locuteur modifie accidentellement la forme, l’usage ou le sens standard d’un phrasème. Sont ensuite étudiées les causes principales de ces erreurs, telles que les troubles langagiers, la faible maîtrise de la langue ou le manque de concentration. Les notions de variation et d’erreur sont alors différenciées lors de l’étude d’exemples problématiques à l’aide d’une combinaison de critères tels que la fréquence proportionnelle, la cohérence sémantique ou l’efficacité communicationnelle. Malgré leur caractère erroné, il est indéniable que certaines combinaisons subissent une lexicalisation et se propagent en raison de divers biais cognitifs. C’est à travers le processus de phraséogenèse que les réponses à ce phénomène a priori paradoxal sont apportées, et plus précisément les critères de sélection qui influencent la dissémination ou « répliation » des phrasèmes.