

Formal and Semantic Features of Creative Phrasal Verbs in Native Speakers' Personal Storytelling: A Corpus-Driven Study

Ruiyuan Wang

Catholic University of Louvain, Louvain-la-Neuve, Belgium

ruiyuan.wang@student.uclouvain.be

Abstract. Most research on linguistic creativity has focused on learners' language use, and creative language use by native speakers, especially in the area of phrasal verbs, has been relatively unexplored. Corpus-driven study of creative phrasal verbs in personal storytelling and an operational definition of creative phrasal verbs: verb-particle constructions that are both unattested in a reference corpus and unarchived in existing phrasal verb dictionaries. A self-compiled corpus of personal stories from The Moth was used to extract phrasal verbs via Corpus Query Language (CQL) searches and then further identified by automated comparison in R. The results show that creative phrasal verbs occur less frequently in personal stories. Most examples are intransitive, particle-inseparable (i.e., VPO/VP structure), and structurally simple (i.e., non-phrasal-prepositional construction), and are more literal than metaphorical in semantics. Some semantic patterns are the same as those in older studies of traditional phrasal verbs, for example, particle redundancy. The study has the following deficiencies: a small corpus size, a descriptive design, a single-rater pattern, an issue with a less comparable reference corpus, and limited extraction criteria. In the future, a large number of various corpora will be used to conduct cross-register comparisons and incorporate diachronic corpus updates for an all-encompassing study of creative language use. Besides, a linguistic comparison between the use of creative and existing phrasal verbs by native speakers, and the statistical predictive power of factors on the semantic findings of this study will be conducted and investigated in the future.

Keywords: phrasal verbs, linguistic creativity, personal storytelling, corpus linguistics, native speakers

1. Introduction

Language users are constantly demonstrating creativity in their daily language use by creating new linguistic forms that effectively convey meaning in particular situations [1-3]. Creativity in language has been recognized for a long time by scholars. Chomsky [4] believes that language users have the ability to create and understand an infinite number of new expressions, and functional approaches focus on how speakers can use linguistic resources for communication [2]. Such creativity is not only in literary and artistic languages but also in daily life.

Although many studies have been conducted on linguistic creativity, most of them have focused on learners' language and second-language production. Compared with other parts of English, creative language use by native speakers has received comparatively little attention. Phrasal verbs are some of the most productive structures in English that have been expanded in terms of meaning and vocabulary [5]. Studies that have examined newly coined or improvised phrasal verbs in authentic native-speaker discourse are also relatively scarce.

Therefore, to address the above problem, the present study will use a corpus-based approach to investigate creative phrasal verb use in American English personal narratives. Personal stories are good materials for studying linguistic creativity because, in speaking about their own lives and feelings, people often use expressive and original language on either a rehearsed or an improvising basis. Based on natural discourse, this paper studies what new phrasal verbs are created by native speakers in their personal storytelling and examines their formal and semantic features, such as transitivity, structural complexity, particle separability, semantic transparency, particle redundancy and metaphorical extension.

By collecting and analyzing creative phrasal verbs used in natural language, this paper hopes to learn more about the linguistic creativity of native speakers and offer empirical support for the productive prospects of English phrasal verbs. The results may also provide a foundation for future research on the connection between phrasal innovation and other speakers' variables in expression today.

2. Literature review

2.1. Phraseological creativity among speaker groups

Most of the studies on the lexical and phraseological creativity have been devoted to non-native and learner language [3, 6]. Limited proficiency, cross-linguistic influence and the complexity of the L2 system have been found to be factors that can lead to the production of innovative or unconventional expressions by L2 learners [2, 6-9]. Such creativity can also manifest itself not only as erroneous use, but also as deliberate language play for interactional purposes [2, 3]. Heritage speakers have also shown to be very creative in the way they use language compositionally [10, 11].

On the other hand, the creativity of native speakers has been much less studied in the field of phraseology. The imbalance might be linked to the fact that native language production is well facilitated by formulaic sequences that are entrenched and enable fluent communication [12, 13]. These native innovations are therefore difficult to detect using traditional structural and/or frequency analyses [14, 15]. However, native speakers are still able to create context-dependent and pragmatically motivated innovations [2]. As phrases are a continuum from idiomatic phrases to flexible combinations [1, 11], the creativity potential and continuum from literal and aspectual to idiomatic of the productive phrases (e.g., phrasal verbs) needs to be explored.

2.2. Phrasal verbs and phraseological productivity

Phrasal verbs (PVs) and their definition and classification are still theoretically discussed. In this study, PVs are defined according to the structural approach proposed in Quirk et al. [16] as combinations of a lexical verb and an adverbial particle, but excluding prepositional structures based on syntactic diagnostics like placement of noun phrases, interruption pattern, pied-piping test, stress distribution, and so forth [17]. Semantically, the traditions tend to be related to idiomatic phrases [14, 16]. However, from the corpus studies, PVs are now viewed as a continuum of

productivity extending from the transparent literal combination to the highly opaque idiomatic expression [5, 18, 19]. Under this scheme, particles can be used to express meanings of direction, aspect, or event structure that enable speakers to build up context-sensitive meanings and keep connected to the existing linguistic patterns [7, 19].

In addition, PV creativity might be related to conceptual metaphor mechanisms, where particles are considered to represent orientational concepts like up–down and in–out [20]. The metaphorical extension and semantic abstraction of support verbs and particles can help to emerge new phraseological meanings [21, 22]. In certain instances, particles can also be semantically redundant, that is, they do not contribute to the meaning of the verb, or overlap with the meaning of the support verbs (e.g., the particle "on" in the phrasal verb: "continue on") [5, 23].

2.3. Phrasal verbs that are creative – operating words in a different way

Although phraseological creativity, especially constructional creativity, has been gaining growing popularity, it is still methodologically tough to identify PV creativity, as there are no golden criteria which define creative phrasal verbs in any existing studies. Thus, this study proposes and is based on a double filtering approach. A creative PV is absent from a representative reference corpus under the same genre or register (or at least fewer than five occurrences in case of contingency) and appears in none of the major lexicographical resources. In other words, phrasal verbs that are considered to be creative are either undiscovered, featured by their scarce occurrences under a certain threshold of minimum frequency, or unarchived. This mixture of frequency analysis based on the corpus and verification in the dictionary allows one to identify the expressions that are novel and to avoid conventional but rare PVs [24].

3. The method and corpus annotation

This study uses as the basis a self-compiled corpus of automatically transcribed personal storytelling excerpts from The Moth, a website that focuses on recording and broadcasting personal stories told by native speakers of American English (<https://themoth.org>). There are 61 text files and 128, 615 words, which are converted to UTF-8 plain-text format and processed in Sketch Engine. Sketch Engine was chosen because it incorporates an automatic parser that produces part-of-speech tags which makes it convenient to carry out a Corpus Query Language (CQL) search to extract phrasal verbs. For sociolinguistic stratification, there are no metadata available because of privacy concerns and the anonymity of the storytellers; this limits the extent to which variables related to speakers can be examined as they relate to creative phrasing of PVs. The storytelling corpus was compared with the spoken part of BNC Corpus 2014 as the reference corpus so as to identify creative uses of phrasal verbs. While COCA (Corpus of Contemporary American English) could have been preferable for data on spoken American English, there were technical limitations and difficulties in accessibility for systematic comparison of the data in the corpus. Therefore, the spoken BNC 2014 was chosen as an alternative accessible expedient because of its size and appropriate register of native-speakers' spoken data. Any phrasal verbs which occurred in the storytelling corpus more than 5 times and which did not appear in the reference corpus were considered as possible creative forms and later checked against the available phrasal verb dictionaries.

A combination of CQL search and manual validation was used for the extraction of phrasal verbs. Part-of-Speech patterns in Sketch Engine were followed and lexical verbs were identified by the tag VV.* and particles by RP.* Multiple query patterns were used to collect both intransitive and transitive phrasal verbs as well as those with inserted objects between the verb and particle. Other

wildcard positions have been added between the verb and particle position to extract constructions that involve pronouns or longer noun phrases as particle-separable phrasal verbs. The procedure allowed to extract phrasal verbs both transitive and intransitive, and with or without intervening constituents.

The identification of the undiscovered type of creative phrasal verbs was further enhanced by an automated comparison procedure in R, which was based on the general looping logic taken from Paquot [11]. The two corpora (storytelling and reference) were then checked for the occurrence of phrasal verbs which were classified as conventional if they occurred in the reference corpus or as creative if they did not. A minimum frequency threshold of five occurrences was used to minimize the effect of forms with low frequencies in the reference corpus, which might be considered as rare, accidental, or likely creative in the reference corpus itself. Then, all the possible creative phrasal verbs were manually verified by the current phrasal verb dictionaries for their non-conventionality.

A number of formal and semantic features of the creative phrasal verbs were manually annotated. The formal features were coded as binary variables: transitivity, particle separability and construction type (single PVs as opposed to PPVs). Semantic annotation aimed at transparency (literal, aspectual or idiomatic) and the existence of metaphorical extension and particle redundancy. All the annotations were done manually based on the context and the definitions set in the literature review. The annotated data is then used in the quantitative and qualitative analyses shown in the sections below and provided as supplementary data for review.

4. Results and analysis

4.1. An overall frequency of the creative phrasal verbs identified in the MOTH corpus

Overall, the paper retrieves a total number of 1304 concordances of phrasal verbs observed from the MOTH corpus. Amidst these concordances, there are 139 phrasal verbs found to be absent from the BNC 2014 (spoken part) reference corpus, and only 52 of them are proven to be unarchived (i.e., not found in the phrasal verb dictionary). The overall result demonstrates a low frequency of the coinage of new phrasal verbs in personal storytelling, with a proportion of less than 4%. The summary of the overview is labelled as follows (Table 1):

Table 1. The absolute frequency and percentage of the total and filtered phrasal verbs

Category	Absolute Frequency	Percentage (%)
Total Phrasal Verbs	1304	100
Absent in Reference Corpus	139	10.66
Unarchived (coined) Phrasal Verbs	52	3.99

4.2. Formal and syntactic features of the creative phrasal verbs found in the MOTH corpus

According to the 52 retrieved and filtered creative phrasal verbs, 43 phrasal verbs do not exhibit any verb-particle separability, which are shown to have support verbs and adverbial particles collocate and co-occur consecutively. This is mainly due to the frequent occurrences of intransitive creative phrasal verbs that do not have an object to interrupt the verb and the particle. However, it is still noteworthy that there are overall 9 creative cases where particle displacement occurs among transitive phrasal verbs, be the objects are one slot or two slots fronted. Among those fronted cases,

6 of them are one-slot fronted where 5 of the fronted items are pronouns (e.g., jam it in), while 3 of the cases have 2 slots-fronted, comprised exclusively of determiner + noun constructions (e.g., sweat my ass off). The distribution and frequency of the particle displacement are illustrated by the bar plot, shown as Figure 1:

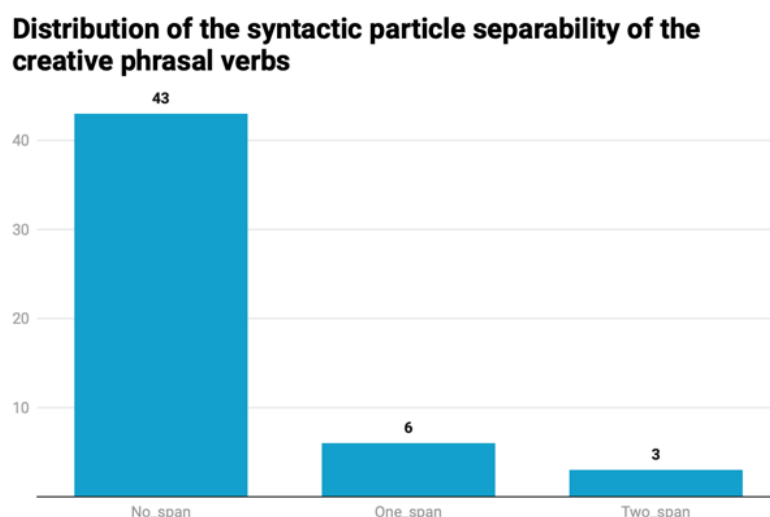


Figure 1. Distribution of the particle-separability of the creative phrasal verbs

This study is also interested in observing the frequency of transitive and intransitive phrasal verbs when used creatively. According to the definitions and the annotation of the corpus, a total number of 17 out of 52 creative phrasal verbs are observed to be transitive, 9 of which are the above-mentioned particle-separable phrasal verbs, as the evidence of fronted objects are already transitive in nature. Therefore, the paper prioritizes the findings on the transitivity of those particle-inseparable verbs. Among the 43 particle-inseparable creative phrasal verbs, only 8 are transitive (18.6%), while 35 are intransitive (81.4%). This result might indicate that the creative particle-inseparable phrasal verbs are less likely to be transitive. The graph that illustrates this finding is displayed as follows:



Figure 2. Proportional frequency of the transitivity of the non-particle-separable phrasal verbs

To finalize the study of the formal and syntactic features of the creative phrasal verbs, the frequency of the simple phrasal verbs (support verb + adverbial particle) and complex phrasal verbs (phrasal-prepositional construction) is extremely interesting (Figure 3). Following the given definition of phrasal-prepositional construction, although most observed phrasal verbs are followed by a preposition, the syntactical and grammatical evidence does not conform to the single-word feature, which is necessary to be considered as a complex phrasal verb, but rather simply functions as a prepositional phrase. Therefore, the cases that have been taken into account are rare, with only 6 occurrences compared with the 46 simple phrasal verbs. The visualization of this feature is displayed as follows:

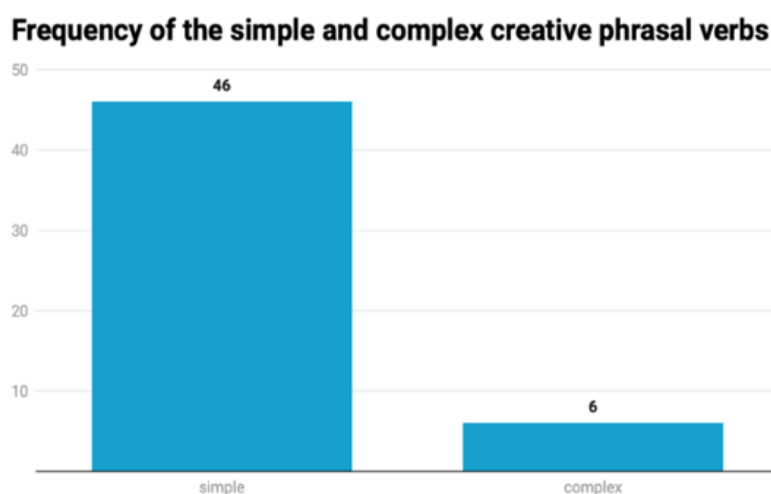


Figure 3. Frequency of the simple and complex creative phrasal verbs

To sum up, from the perspective of syntax and form, creative phrasal verbs discovered in the personal storytelling context exhibit less particle-separable syndrome, dominantly coined for intransitive use, and are mostly simple and typical-like (Figure 3).

4.3. Semantic features of the creative phrasal verbs found in the MOTH corpus

The stratification of the semantic transparency on a continuum is previously mentioned and is applicable to the creativity of the phrasal verbs. Amidst the 52 concordances, a majority of 37 cases are annotated as semantically literal, and 15 cases are marked either aspectual (10) or idiomatic (5). The finding illustrates a dominance of semantically transparent (i.e., literal) phrasal verbs in the coinage, while the opaque idiomatic ones and functionally aspectual ones are less primed in the actual use of storytelling.

As for the stratification of the particle redundancy, the results have found a relatively average distribution of phrasal verbs with redundant particles and indispensable particles, where 22 cases display compulsory verb-particle collocation while 30 cases where particles that are unnecessary. This result does not illustrate a significant general tendency of the necessity of adverbial particles in coined phrasal verbs.

However, an interesting converging finding may reveal the association between the semantic transparency and particle redundancy of the creative phrasal verbs, as the 9 cases for the aspectual category are found to have completely redundant particles while the 5 occurrences of idiomatic phrasal verbs are extremely reliant on indispensable particles. In other words, the 9 aspectual phrasal verbs are marked as redundant in terms of the particles, and the 5 idiomatic phrasal verbs turn out to be just the opposite. This finding is coherent with and empirically corroborates the featured definitions of the aspectual and idiomatic phrasal verbs. It is only the elements in the literal category that collocate with both redundant and indispensable particles. Among the 37 literal phrasal verbs, 16 of them are indispensable, while the redundant phrasal verbs of this category are slightly more, with 21 occurrences marked as unnecessary, which shows a seemingly average distribution of their redundancy (figure 4). The mosaic plot that substantiates this finding is shown as follows:

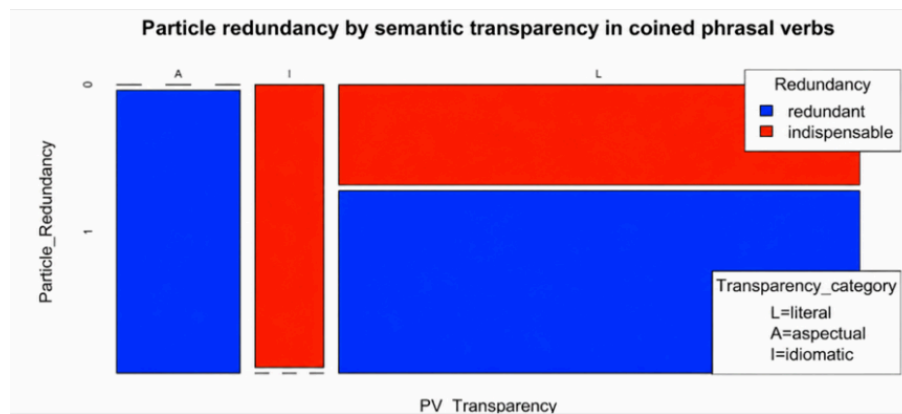


Figure 4. Proportional distribution of particle redundancy across transparency categories

Knowing that the phrasal verbs are metaphoric, this study also discovers the metaphoric patterns of the creative phrasal verbs. Concerning the definitions of the metaphoric phrasal verbs, there are only 7 out of a total of 52 cases that are found to be metaphoric, while the 45 cases are non-metaphoric. At this phase, the paper finds that the metaphoric creative phrasal verbs are quite rare. Among the 7 cases of the metaphoric ones, only 2 phrasal verbs display orientational metaphor on particles, while the majority of 5 are found to carry metaphor mainly on support verbs, nevertheless. Interestingly, although statistical substantiation by a larger dataset has not yet been conducted, the separable creative phrasal verbs feature a metaphor reliance on adverbial particles, whereas the inseparable ones carry more metaphoric meanings on their support verbs. As for the association between the transparency category and the metaphoric use, all 5 idiomatic phrasal verbs are metaphoric, whereas there are 2 cases of literal phrasal verbs that are metaphoric as well, which is intriguing. However, due to the small sample, this paper does not attempt any generalization of these observed findings (table2). The stated results are presented visually as follows:

Table 2. An overview of metaphoric categories of identified creative phrasal verbs

Metaphor Categories	Raw Frenquency	Percentage
Non-metaphorical	45	86.54
Verb Metaphor	5	9.62
Particle Metaphor	2	3.84
Total	52	100

Association between the frequencies of metaphoric patterns and transparency categories



Figure 5. Association between the metaphoric patterns and transparency of the metaphoric phrasal verbs

To summarize the findings, creative phrasal verbs are observed to be relatively rare in the personal storytelling context. As for the syntactic and formal features, creative phrasal verbs are

more likely to be intransitive, which indirectly complements the findings in terms of the less frequent cases of particle-separable structure. Besides, the creative phrasal verbs are largely found to be typical and simple, with adverbial particles following the support verbs directly, while particle-prepositional constructions are rare. As for the semantic features, creative phrasal verbs exhibit a strong tendency to be literal, but less aspectual and idiomatic. In regard to the particle redundancy, creative phrasal verbs do not show any preference for either. However, except for the literal phrasal verbs whose distribution of redundancy is relatively average, the other two categories display extreme binary predilection of particle indispensability and redundancy. Moreover, creative phrasal verbs are used less for metaphoric purposes. However, non-particle-separable phrasal verbs and idiomatic phrasal verbs are likely to be more "verbally" metaphorical, although this finding is not statistically significant(Figure 5).

5. Discussion

Based on the above results, it can be inferred that creative phrasal verbs are relatively infrequent in personal narratives, possibly due to prior preparation for presentation. However, the appearance of creative forms also suggests that lexical innovation can be an intentional stylistic device employed to increase expressiveness and appeal to the audience.

In terms of formal features, non-particle-separable phrasal verbs occurred at a much higher rate than particle-separable ones. Although this is not the same as Gilquin's, the difference may be due to the selection criteria for extraction in the present study, which excluded extended NP insertions [25]. Semantically, literal phrasal verbs made up the majority of the dataset; therefore, transparent constructions may be more suitable for spoken narration due to their ease of processing and comprehension. The low frequency of metaphorical creative phrasal verbs and a small corpus size hinder further generalization. Future research will use a larger dataset to explore more formally and semantically the patterns of creative phrasal verbs in personal stories.

This corpus-driven study aims to fill a deficiency in the research of native speakers' linguistic creativity at both the level of definition and method. Although it adheres to the traditional mainstream syntactic criteria, this study considers a broader range of semantic spaces for verb-particle constructions and, in its original work, defined creative phrasal verbs (PVs) as forms that are both unused and unarchived, establishing a looping extraction and filtering protocol to identify them. Based on the empirical results, phraseological innovation is relatively infrequent in the context of personal storytelling, and creative PVs account for less than 4% (52 out of 1,304 concordances) of the total occurrences. Structurally, most of these creative coinages are simple and typical; they are mainly intransitive, have little particle separability, and rarely use complex phrasal-prepositional constructions. Semantically, literal and transparent forms make up most of the data; that is to say, easily processed structures are preferred to reduce the cognitive load of spoken communication, and some categories show a strong binary preference for particle redundancy. There are some defects in the breadth and methods of this study. It is purely descriptive and has a small size (128,615 words); therefore, robust inferential statistical analysis cannot be performed on this sample to explore any relationships. In addition, extended object insertions beyond the two filler slots may have neglected relevant phrasal verb concordances that have longer NPs between verbs and particles, and due to privacy reasons, the storyteller's metadata was unavailable for sociolinguistic stratification. Given the risk of subjective bias in single-rater manual annotation, it is more reasonable to add an inter-rater reliability test, such as Cohen's Kappa, to the design of the next study. In addition, the baseline comparison had technical and dialectal limitations; using the spoken British National Corpus instead of the American COCA introduced regional mismatches, and some

"creative" forms that were not in BNC 2014 were found to appear in newer web corpora (i.e., TenTen series corpora). To overcome the above limitations, future research should increase the size of the corpus, use dialect-matched reference databases, employ multi-rater validation, and introduce cross-sectional or diachronic comparisons of registers and speaker groups to provide a more complete and detailed study of native linguistic creativity. Besides, some findings of this research should be correlated and associated by inferential statistics to make the results more coherent. For example, with regard to the semantically literal phrasal verbs, this study also tends to statistically investigate the factor(s) that predict particle redundancy of creative phrasal verbs to compare with the predictive factors of particle redundancy of existing phrasal verbs. The further research as such would reveal the reasons behind native speakers' seemingly "redundant" language creation and the cognitive patterns that monitor their language production.

6. Conclusion

The corpus-based study of creative phrasal verbs used in American English personal narratives combines corpus extraction, reference corpus comparison, and manual annotation. 52 unarchived creative phrasal verbs in the Moth storytelling corpus were identified. Based on the above results, creative phrasal verbs are relatively rare and account for less than 4% of all observed phrasal verb occurrences. However, these cases show that native speakers may take advantage of rehearsals to create context-specific expressions.

The results show that most creative phrasal verbs are in the conventional structure. They are mainly simple verb-particle combinations, and they prefer intransitive use and have limited particle-separable structures. Semantically, creative phrasal verbs are generally literal and transparent; therefore, easily understandable expressions may be preferred in personal narration. Although metaphorical and idiomatic innovations have been identified, they are relatively infrequent and require further study.

This study has some deficiencies. A small corpus limits the generalizability of the results and cannot be used for further statistical tests. There is no speaker metadata available for the analysis of how speakers' related variables affect the use of creative language. A reference corpus of British English may have been used to identify creative forms due to the accessibility issues of COCA corpus, which therefore introduces dialectal differences. Future work will increase the size of the corpus, use appropriate American English reference materials, and add multiple annotators to enhance the reliability of classification. Comparative studies in different registers, speaker groups and at various times have also shown how phrasal creativity evolves and how native speakers continuously innovate language use.

References

- [1] Cowie, A. P. (2014). Stable and creative aspects of vocabulary use. In *Vocabulary and language teaching* (pp. 126-139). Routledge.
- [2] Kecskes, I. (2016). Deliberate creativity and formulaic language use. *Pragmemes and theories of language use*, 3-20.
- [3] Bell, N. (2012). Formulaic language, creativity, and language play in a second language. *Annual Review of Applied Linguistics*, 32, 189-205.
- [4] Chomsky, N. (2011). *Current issues in linguistic theory (Vol. 38)*. Walter de Gruyter.
- [5] Bolinger, D. L. M. (1971). *The phrasal verb in English*. Harvard University Press.
- [6] Riguel, E. (2014). Phrasal verbs, "the scourge of the learner". In *9th Lancaster University Postgraduate Conference in Linguistics & Language Teaching (Vol. 9)*, pp. 1-20.

- [7] Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, 32, 130-149.
- [8] Kecskes, I. (2007). Formulaic language in English lingua franca. *Explorations in pragmatics: Linguistic, cognitive and intercultural aspects*, 1, 191-218.
- [9] Kecskes, I. (2015). Is the idiom principle blocked in bilingual L2 production? *Bilingual figurative language processing*, 28-53.
- [10] Rakhilina, E., Vyrenkova, A., & Polinsky, M. (2016). Linguistic creativity in heritage speakers. *Glossa: A Journal of General Linguistics*, 1(1), 1-27.
- [11] Paquot, M. (2019). The phraseological dimension in interlanguage complexity research. *Second Language Research*, 35(1), 121-145.
- [12] Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combinations. *Phraseology: Theory, analysis, and applications*, 101-123.
- [13] Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *Language and communication*, 191-226.
- [14] Howarth, P. A. (2013). *Phraseology in English academic writing: Some implications for language learning and dictionary making (Vol. 75)*. Walter de Gruyter.
- [15] Granger, S., & Paquot, M. (2008). Disentangling the phraseological web. In *Phraseology: An interdisciplinary perspective* (pp. 27-49). John Benjamins Publishing Company.
- [16] Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Longman.
- [17] Darwin, C. M., & Gray, L. S. (1999). Going after the phrasal verb: An alternative approach to classification. *TESOL Quarterly*, 33(1), 65-83.
- [18] Moon, R. (1998). *Fixed expressions and idioms in English: A corpus-based approach*. Oxford University Press.
- [19] Armstrong, K. (2004). Sexing up the dossier: A semantic analysis of phrasal verbs for language teachers. *Language Awareness*, 13(4), 213-224.
- [20] Lakoff, G., & Johnson, M. (2024). *Metaphors we live by*. University of Chicago Press.
- [21] Tsaroucha, E. (2020). Embodiment and image schemas: Interpreting the figurative meanings of English phrasal verbs. *Languages*, 5(1), 6.
- [22] Garnier, M. (2016). *English phrasal verbs: Usage, knowledge, acquisition* (Doctoral dissertation, University of Nottingham).
- [23] Lipka, L. (1972). *Semantic structure and word-formation: Verb-particle constructions in contemporary English*. Wilhelm Fink Verlag.
- [24] Chomsky, N. (1981). Principles and parameters in syntactic theory. *Explanation in Linguistics*, 32-75.
- [25] Gilquin, G. (2025). Second and foreign language learners: The effect of language exposure on the use of English phrasal verbs. *International Journal of Bilingualism*, 29(2), 456-473.