

ANTONINA BONDARENKO

ORCID ID: <https://orcid.org/0000-0003-3622-296X>

Paris Diderot University

Predication Transformation: A Parallel Corpus-Based Study of Russian Verbless Sentences and Their English Translations

Transformacja predykacyjna: oparte na równoległych korpusach studium
rosyjskich zdań bezczasownikowych i ich angielskich tłumaczeń

1. INTRODUCTION

Based on a pilot parallel corpus, this paper examines the way that predication is rendered in the translation of verbless sentences from Russian to English. The definitions of predication and predicate, and the verb's role in these notions, have been highly debated in linguistics. Special issues of *Faits de Langues: La Prédication* (2009) and *Revue de Linguistique et de Didactique des Langues: Syntaxe et Sémantique des Prédicats* (2008) trace the notions throughout history. In this paper, we apply Kees Hengeveld's (1992) definitions to a series of verbless sentences and their translations which have been automatically extracted from a small literary parallel corpus. Following Jaqueline Guillemin-Flescher's (2003) approach to contrastive linguistics, we explore re-occurring patterns with regard to the way that predication is gained or lost in translation.

The comparison of Russian verbless sentences with English translations is particularly relevant due to the profound cross-linguistic differences between the two languages. Sentences in which the grammatical category of verb is absent exist in many languages, but they are known for being particularly frequent in Russian. Out of all Indo-European languages, Russian is famous for allowing the most liberal use of verbless sentences (Kopotev, 2007b). Characteristic features of Russian include a very developed morphological case system, flexibility of word

order and intonation, absence of articles, an extraordinary capacity for verbal ellipsis, the possibility of a zero-copula construction in the present tense – such as (1) “Я Алексей” (*ja Alexej*, lit. ‘I Alexei’), and many other non-elliptical verbless constructions – such as (2) “Я в монастырь” (*ja v monastyr’*, lit. ‘I to monastery’) uttered in a context without a linguistically explicit verbal antecedent (Stassen, 2013). In contrast, English is known for its dependence on the finite verb phrase, the lack of a zero-copula construction and formal register restrictions on certain types of verbal ellipsis (McShane, 2000). Nonetheless, verbless sentences are also found in English across all sentences types, including typical exclamatives, e.g. (3) “What a picture!”, imperatives, e.g. (4) “And now to business.”, questions, e.g. (5) “What about my parental blessing?”, and assertions, e.g. (6) “So much, then, for the introduction”.

The paper starts with a description of the corpus and the methodology which involves the automatic extraction of verbless sentences. In Part 4, we discuss the definitions used in automatic and manual annotation. This section includes a sketch of Hengeveld’s (1992) semantic notion of verbal and non-verbal predication, which we develop to include a distinction for antecedent-based ellipsis. The results in the final part present descriptive statistics with particular attention to the sentences that gain a verb in translation, characterize the phenomenon of predication transformation according to five different types, and analyze the implication of the observed translation patterns for the semantic notion of predication. We conclude by summarizing the limits of the current pilot study and suggesting perspectives for further research.

2. PARALLEL CORPUS

The corpus consists of Fyodor Dostoyevsky’s Russian dialogue-centered *Братья Карамазовы* (*Brat’ja Karamazovy*, 1880) and the Richard Peaver and Larissa Volokhonsky English translation *The Brothers Karamazov* (1990). This Russian novel has inspired many studies in literature and philosophy, but Dostoyevsky’s language has also been praised as particularly suitable for the study of spoken dialogue. George Thomas (1982: 672) stresses that the important role Dostoyevsky gives to dialogue makes this novel of particular interest for linguists, especially those studying speech acts. Targeting the key features of reliable parallel corpora described by Thomas Stolz (2007), we selected this work due to the frequent passages of direct speech, everyday language register of the original, realistic prose, and the existence of sixteen translations of the novel. The translation was chosen for its recency, basis on the original text, and its literal style (Vasil’čenko, 2007)

which has earned it critical acclaim for being most true to the original. Although for reasons of feasibility of the pilot study we examined only one translation, the large number of competing translations make it possible to compare the patterns across multiple different translators in future work.

The scope of the corpus is limited to the first fourteen chapters of the novel and contains a total of 76,500 words. The manageable size allowed us to develop a new approach to extracting verbless sentences automatically, verify the accuracy of the extraction, and manually annotate the verbless sentences and their translation correspondences in accordance with the definitions of predication and the categories described in Part 4.

3. VERBLESS SENTENCE EXTRACTION

While much progress has been made in natural language processing with regard to the search for a particular word or element in a corpus, finding the absence of an element in a sentence automatically still remains a challenge. Studies of existing parsed corpora show that very often verbless sentence extraction is blocked by syntactic modeling that is based on verb-centric definitions of a clause and the typically fixed morphosyntactic annotation (Landolfi et al., 2010).

We try to resolve the challenges of automatic processing of verbless sentences by developing an alternative method of automatic extraction. By customizing automatic sentence segmentation, semi-automatically correcting morphosyntactic annotation by means of Trameur annotation, alignment and statistical text analysis software (Fleury and Zimina, 2014), and using the latter to classify the sentences into those with a verb and those without, we achieved an accuracy of 94% in terms of automatic recall of verbless sentences as compared with manually extracted results.

4. ANNOTATION DEFINITIONS

Automatically extracted verbless sentences include any structure that ends with a major punctuation mark and does not show a verb, or verb form (participles, infinitives), in any of its parts. Direct speech sentences were separated from embedded narration in automatic segmentation. Following extraction, verbless sentences were aligned by paragraph in order to visualize the context, they were separated into utterances, and each verbless utterance and its translation were manually annotated for the presence of a verb, the presence of an antecedent-based

verbal ellipsis following Marjorie McShane (2000) and Mikhail Kopotev (2007a), and whether the predication involved was verbal or non-verbal in accordance with Hengeveld (1992).

A key element of Hengeveld's (1992) definitions is that the notion of non-verbal predication is wider than the notion of a verbless sentence. He stresses that the notion of non-verbal predication is a semantic notion that may be morphosyntactically expressed by both verbless and verbal sentences. Non-verbal predication is defined as taking place in all constructions where a non-verbal predicate is applied to arguments (Hengeveld, 1992: 26). The non-verbal predicate "should be considered the main predicate of a non-verbal predication, even in those cases in which it is accompanied by a copula" (Hengeveld, 1992: 26). Therefore, on Hengeveld's conception, sentences containing a verbal, but semantically empty, copula are treated as instances of non-verbal predication. These distinctions are illustrated in Figure 1, which is based on Hengeveld's (1992: 27) diagram and includes a modification for ellipsis and examples from the present corpus.

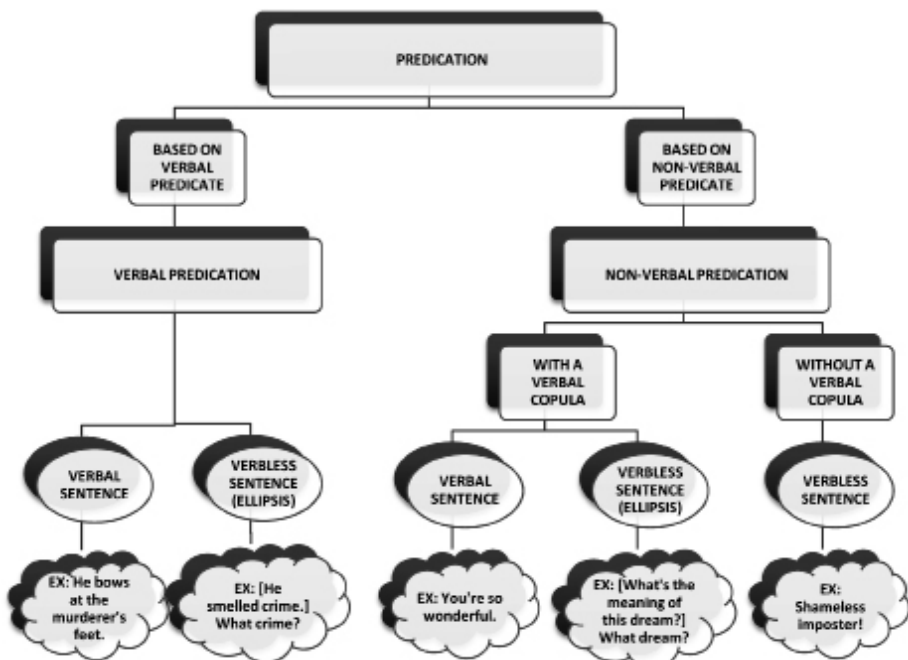


Figure 1. Predication and sentence distinctions based on Hengeveld (1992), extended to include antecedent-based ellipsis and illustrated with examples from the present corpus

Source: author's own elaboration

In the same vein as Emile Benveniste (1966: 163) and Rodney Huddleston and Geoffrey Pullum (2002: 218), Hengeveld (1992) treats the copula verb “be” as semantically empty. However, he argues that the semantically empty copula cannot constitute the main predicate of the sentence. Therefore, in Hengeveld’s (1992) terms, the main predicate in example (7) would be “so wonderful”, not “is” or “is so wonderful”, and the English verbal sentence would be treated as a case of non-verbal predication.

(7) [– Ах, как это с вашей стороны мило и великолепно будет, – вдруг, вся одушеваясь, вскричала Lise. – А я ведь маме говорю: ни за что он не пойдет, он спасается.]

[“Ah, how nice and splendid it will be of you,” Lise cried with sudden animation. “And I just said to mother: he won’t go for anything, he is saving his soul.”]

Экой,	экой	вы	прекрасный!
PRO. M. S. NOM	PRO. PRO. M. S. NOM	2PL. NOM	ADJ. M. S. NOM
what-a	what-a	you	wonderful
“You’re so wonderful, so wonderful!”			

In applying Hengeveld’s (1992) definitions, we include antecedent-based ellipsis. Example (8) illustrates sluicing ellipsis in both languages, following McShane’s (2000) discussion of ellipsis types and licensing conditions, and demonstrates predication transformation.

(8) [– Скажи ты мне, Алексей одно что сей сон значит? Я вот что хотел спросить.]

[“Tell me one thing, Alexei: what’s the meaning of this dream? That’s what I wanted to ask you.”]

Какой	сон?
PRO. M. S. NOM	M. S. NOM
what	dream
“What dream?”	

Since the ellipted antecedent is the semantically meaningful finite verb “значит” (*značit*, lit. ‘means’), the predication involved in the Russian verbless sentence is annotated verbal. In English, the antecedent found in the previous clause is the finite form of the copula verb “be” and the predication is annotated non-verbal. Therefore, a transformation from verbal Russian predication to non-verbal English predication has occurred. The results in the following part analyze the predication transformation phenomenon from a quantitative and qualitative perspective.

5. PREDICATION TRANSFORMATION

5.1. Descriptive statistics

Automatic extraction revealed 315 Russian verbless sentences out of a total of 2,325 sentences, therefore, establishing a *verbless sentence rate* of 13.5% for the Russian corpus. Since a sentence may consist of several utterances and contain several verbs, the manual annotation was performed on utterances. A total of 419 Russian verbless utterances and their translations were examined. The translation results show an *utterance verbalization rate* of 49%, that is 207 of the 419 Russian verbless utterances gained a verb in English translation.

The loss or gain of a finite verb in the translation of an utterance often resulted in a change of the predication type of the utterance. While the *utterance verbalization rate* concerns the change in the realization of the grammatical category of the verb, the *predication verbalization rate* addresses the change in the predication type involved in the verbless utterances as compared to their translations. The definitions summarized in Figure 1 above were used to determine the type of predication, i.e. (a) *verbal predication*, involving a non-copula verb (or a non-copula verbal antecedent), or (b) *non-verbal predication*, which either involves a copula verb (or copula verb antecedent) or involves no verb (or verbal antecedent) at all.

Transformation from Russian verbal predication to English non-verbal predication, such as (8) above, was a rare phenomenon. Only six cases were identified, all of which involved an ellipsis of the copula “be” in English. Predication transformation essentially occurred in the opposite direction, that is from Russian non-verbal to English verbal. *Predication verbalization*, the result of the gain of a non-copula verb in the translation of a non-elliptical verbless utterance, was observed in 19% of the English translations of Russian verbless utterances. The verbs involved in predication transformation include: *bow, can, care, come, damn, serve, do, drive, fast, follow, get, go, happen, have, hold, know, leave, like, look like, make, mean, need, say, think, treat, want*.

5.2. Typology

A closer examination of the cases where non-verbal predication became verbal in translation from Russian to English reveals five patterns in which the transformation occurred.

Idiomatic expressions

The first involves the introduction of verbs as part of the translation of idiomatic expressions. These are fixed expressions that are semantically non-compositional, that is their lexical elements do not make transparent the meaning of the sentence transparent (Kopotev, 2015: 226). For instance, the gain of the verb, and the consequent predication transformation, in the exclamative in (9) is explained by the fact that this is a fixed Russian expression for which the translator resorts to a fixed English expression in order to scold the interlocutor for his actions.

(9) [– Чего же ты? Зачем ты его так? – вскинулся Федор Павлович, но коляска уже поехала. Иван Федорович не ответил. – Ишь ведь ты! – помолчав две минуты, проговорил опять Федор Павлович, косясь на сына: – сам ведь ты весь этот монастырь затеял, сам подстрекал, сам одобрял, чего ж теперь сердиться?]

[“What’s got into you? Why did you do that to him?” Fyodor Pavlovich heaved himself up, but the carriage was already moving. Ivan Fyodorovich did not answer. “How do you like that?” Fyodor Pavlovich said again after two minutes of silence, looking askance at his boy. “You started this whole monastery business, you urged it, you approved it, why are you angry now?”]

Ишь	ведь	ты!
PART	PART	2S.NOM
oh	after-all	you
“How do you like that?”		

Although the introduction of the verb “like” in translation transforms the predication type, its lexical status is uncertain due to the idiomatization of the English expression of which it is a part. Moreover, the fact that the English utterance is a rhetorical question, used to indirectly assert disapproval of the previous turn, further minimizes the verb’s semantic contribution to this instance of predication.

Untranslatable part of speech

In other cases, verbs were used in the translation of words that do not seem to have a non-verbal semantic equivalent in the target language. In (10) non-verbal predication becomes verbal in translation due to the fact that a single noun for “people who fast” does not seem to exist in English colloquial register.

(10) [Ну, а здесь ничего, здесь нет монастырских жен, а монахов штук двести. Честно. Постники. Сознаюсь...Гм.]

[Well, there’s nothing like that here, no monastery wives, and about two hundred monks. It’s honest. They fast. I admit it...H’m.]

Постники.
 M.PL.NOM
 people-who-fast
 “They fast.”

The lexical verb is used in the English sentence in order to preserve the semantic import content of the Russian noun.

Emphasis

The third predication transformation pattern concerns emphasis, as illustrated in (11).

(11) [Знаешь ты штуку?]
 [I'll tell you one thing:]

Пусть	он	и	честный	человек,	Митенька-то
CONJ3S.M.NOM		CONJ	ADJ.M.S.NOM	M.S.NOM	M.S.NOM-PART
though	he	and	honest	person,	Mitenka-that

(он	глуп,	но	честен);	но	он	–
3S.M.NOM	ADJS.M.S	CONJ	ADJS.M.S	CONJ	3S.M.NOM	
(he	stupid	but	honest);	but	he	–

сластолюбивик.
 M.S.NOM
 sensualist

“Granted he’s an honest man, Mitenka, I mean (he’s stupid but honest), still he’s a sensualist.”

The communicative function of the verbal predication in the utterance “I mean” is to emphasize “*Mitenka*”, the antecedent of the pronoun “he”. The corresponding focus is created in the Russian utterance by means of the particle “-то” (*-to*, lit. ‘that’) added onto the proper noun. The introduction of the verb, resulting in predication transformation, is a means of creating emphasis and keeping the information structure intact.

The topic-subject

In other instances, predication transformation is associated with evoking the subject for topic activation. This pattern is illustrated in (12), where in the English translation both a subject and a verbal predicate are gained as compared with the Russian source.

(12){The woman tells a story about herself. The elder asks her a question.}
 Издалека?
 ADV
 from-far-away
 “Have you come from far away?”

The introduction of the English subject pronoun “you” makes the topic of the sentence linguistically-explicit, which in accordance with Knud Lambrecht (1994) is the woman. The activation of the topic by means of evoking the subject results in the introduction of the predication transforming verb in the English utterance.

Contextually implied

Finally, the verbalization of predication in translation of Russian verbless utterances also resulted from the explicit activation of verbs referring to extra-linguistically accessible activities. For example, in (13) only a verb of speaking or meaning would be appropriate for the context.

(13) [– Расславят, запомнят: преступление, дескать, предугадал, преступника отметил. У юродивых и все так: на кабаке крестится, а в храм камнями мечет. Так и твой старец: праведника палкой вон, а убийце в ноги поклон. – Какое преступление? Какому убийце? Что ты? – Алеша стал как вкопанный, остановился и Ракитин.]

[“They’ll proclaim it, they’ll remember: ‘He foresaw the crime and marked the criminal.’ It’s always like that with holy fools: they cross themselves before a tavern and cast stones at the temple. Your elder is just the same: he drives the just man out with a stick and bows at the murderer’s feet.” “What crime? What murderer? What are you saying?” Alyosha stopped dead. Rakitin also stopped.]

Что	ты?
PRO.ACC	2S.NOM
what	you
“What are you saying?”	

By introducing the verb “say”, the English translation linguistically activates the relationship between “what” and “you” which is already contextually accessible. The identical Russian structure may be felicitously used in another context, for example, physical movement (e.g. “jump”) or emotion (e.g. “laugh”). The semantic value of the predication transforming verb in the English sentence seems to overlap with activities that are contextually salient for the interlocutors.

CONCLUSION

The corpus-based analysis of the predication involved in the translation of verbless utterances from Russian to English in accordance with Hengeveld’s (1992)

definitions has revealed a phenomenon of predication transformation. The transformation of predication type occurs when a non-copular verb is gained or lost in the translation of a verbless utterance. The present pilot corpus showed that Russian verbless utterances gained a verb at a rate of in 49% of cases, and that a transformation from Russian non-verbal predication to English verbal predication occurred in 19% of utterances. A typology of the transformation phenomenon suggests that predication transforming verbs are introduced in translation as part of an idiomatic expression, to preserve the semantic import of another part of speech, to create emphasis, as part of topic-subject activation, or to activate a salient element of the extra-linguistic context. As a result, it appears that the semantic contribution of the verb in establishing a division within the semantic notion of predication is questionable. The results of the present study imply that a verbal versus non-verbal dichotomy is inadequate for a cross-linguistically stable definition of the semantic notion of predication.

The parallel corpus of the present study was limited in order to permit the development of an accurate automatic method of verbless sentence extraction and the manual annotation of predication type. A larger and more diverse corpus that is representative of the English and Russian language would make it possible to extend the conclusions beyond the present pilot corpus. Furthermore, the results of the present study need to be verified bi-directionally, that is, in a corpus of Russian translations from an English source. It remains to be seen to what extent such a corpus would confirm or deny the present conclusions concerning the verbs that are gained or lost in translation. We expect that an English to Russian parallel corpus would reveal de-verbalization rates, similar to the present rates of utterance verbalization, and a suppression of non-copular verbs in Russian translations of English verbal utterances in accordance with the presented typology of predication transformation.

REFERENCES

- Benveniste, E. (1971). *Problems in General Linguistics*. (M.E. Meek, Trans.). Miami: University of Miami Press, (original work published 1966).
- Fleury, S. & Zimina, M. (2014). Trameur: A framework for annotated text corpora exploration. In: L. Tounsi & R. Rak (eds.), *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: System Demonstrations, August 2014, Dublin, Ireland* (57–61), <http://www.aclweb.org/anthology/C14-2013.pdf>; hal-01223725 (10.01.2018).
- Guillemain-Flescher, J. (2003). Théoriser la traduction. *Revue française de linguistique appliquée*, 8(2), 7–18.
- Hengeveld, K. (1992). *Non-Verbal Predication: Theory, Typology, Diachrony*. Berlin: Mouton de Gruyter.
- Huddleston, R. & Pullum, G.K. (2002). *The Cambridge Grammar of the English Language*. Cambridge: CUP.

- Kopotev, M. (2007a). What a difference a verb makes! Russian and Finnish verbless sentences. In: M. Nenonen & S. Niemi (eds.), *Collocations and Idioms 1: Papers from the First Nordic Conference on Syntactic Freezes, Joensuu, 20*, 177–192.
- Kopotev, M. (2007b). Where Russian Syntactic Zeros Start: Approaching Finnish? *Slavica Hel-singiensa*, 32, 116–137.
- Kopotev, M. (2015). Reconstruction and Idiomaticity: The origin of Russian verbless clauses recon-sidered. *Folia Linguistica Historica*, 36(1), 219–243.
- Lambrecht, K. (1994). *Information Structure and Sentence Form : Topic, Focus and the Mental Representation of Discourse Referents*. Cambridge: CUP.
- Landolfi, A., Sammarco, C. & Voghera, M. (2010). Verbless clauses in Italian, Spanish and English: A Treebank annotation. In: S. Bolasco, I. Chiari & L. Giuliano (eds.), *JADT 2010: Proceedings of the 10th International Conference on Statistical Analysis of Textual Data* (pp. 1187–1194). Rome: CISU.
- Merle, J.-M. (ed.). (2009). *Faits de Langues: La Prédication*, 31–32, 1–462.
- McShane, M. (2000). Verbal ellipsis in Russian, Polish and Czech. *The Slavic and East European Journal*, 44(2), 195–233.
- Novakova, I. & Guentchéva, Z., (eds.) (2008). *Revue de linguistique et de didactique des langues: Syntaxe et Sémantique des Prédicats*, 37, 1–199.
- Stassen, L. (2013). Zero copula for predicate nominals. In: M.S. Dryer, & M. Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/120> (access: 22.08.2018).
- Stolz, T. (2007). *Harry Potter meets Le Petit Prince*: On the usefulness of parallel corpora in cross-linguistic investigations. *STUF-Sprachtypologie und Universalienforschung*, 60(2), 100–117.
- Thomas, G. (1982). Aspects of the study of Dostoyevsky’s vocabulary. *The Modern Language Review*, 77(3), 670–678.
- Vasil’chenko, T. V. (2007). Sovremennye Anglojazyčnye Perevody Romana F.M. Dostoevskogo “Brat’ja Karamazovy”: K Voprosu o Perevodčeskix Strategijax v Anglojazyčnoj Kul’ture Konca XX v. *Vestnik Tomskogo Gosudarstvennogo Universiteta*, 305 (Dekabr’), 7–10.

ABSTRACT

The paper presents a corpus-based contrastive analysis of the predication involved in the translation of verbless sentences from Russian to English based on a pilot parallel corpus consisting of Dostoyevsky’s dialogue-based Russian *Brat’ja Karamazovy* (1880) and the Pevear and Volokhonsky English translation *The Brothers Karamazov* (1990). In contrast to English, known for its dependency on the finite verb phrase, Russian permits the use of verbless sentences more productively than any other Indo-European language (McShane, 2000; Kopotev, 2007). Combining the parallel-text approach to contrastive linguistics developed by Guillemin-Flescher (2003) with a new method of automatic verbless sentence extraction, the present study examines reoccurring patterns regarding the way that predication is gained or lost in translation. Following automatic segmentation, morphosyntactic annotation and extraction, verbless sentences and their translation correspondences are manually annotated for verbal and non-verbal predication in accordance with Hengeveld’s (1992) definitions. The results present a typology of a phenomenon we call ‘predication transformation’, in which translation correspondences are transformed in terms of predication type. Quantitative results reveal the rate at which verbs are gained in translation of the verbless sentences from Russian to English, as well as the predication verbalization rate. We argue that a verb-centric notion of semantic predication is not cross-linguistically stable.

Keywords: verbless sentences, parallel corpus, predication, English, Russian

ABSTRAKT

Artykuł przedstawia opartą na korpusie, kontrastywną analizę predykcji związanej z tłumaczeniem zdań bezczasownikowych z języka rosyjskiego na język angielski w oparciu o pilotażowy paralelny korpus składający się z opartego na dialogach języka rosyjskiego *Braci Karamazow* Fiodora Dostojewskiego (1880) i angielskiego tłumaczenia Peveara i Volokhonsky *The Brothers Karamazov* (1990). W przeciwieństwie do angielskiego, znanego z zależności od frazy werbalnej z osobową formą czasownika, rosyjski pozwala na użycie zdań bezczasownikowych bardziej produktywnie niż jakkolwiek inny język indoeuropejski (McShane, 2000; Kopotev, 2007). Łącząc założenia tekstu paralelnego w odniesieniu do językoznawstwa kontrastywnego opracowane przez Guillemin-Flescher (2003) z nową metodą automatycznej ekstrakcji zdań bezczasownikowych, niniejsze studium bada powtarzające się wzorce dotyczące sposobu, w jaki struktury językowe zyskują lub tracą wartość predykcji w tłumaczeniu. Po automatycznej segmentacji, morfosyntaktycznej adnotacji i ekstrakcji zdania bezczasownikowych i ich tłumaczeniowe odpowiedniki są ręcznie przypisywane do typu predykcji werbalnej i niewerbalnej zgodnie z definicjami Hengevela (1992). Wyniki przedstawiają typologię zjawiska zwanego “transformacją predykacyjną”, w której odpowiedniki translacyjne są efektem transformacji ze względu na typ predykacyjny. Wyniki ilościowe ukazują, w jakim stopniu czasowniki są wykorzystywane w tłumaczeniu zdań bezczasownikowych z języka rosyjskiego na język angielski, a także są wskaźnikiem werbalizacji predykatów. Twierdzimy, że powiązane z czasownikiem pojęcie semantycznej predykcji nie jest stabilne międzyjęzykowo.

Słowa kluczowe: zdania bezczasownikowe, korpus paralelny, predykcja, język angielski, język rosyjski