Predication Transformation: A parallel corpus-based study of Russian verbless sentences and their English translations
Antonina Bondarenko

To cite this version:

HAL Id: hal-02177744
https://hal.archives-ouvertes.fr/hal-02177744
Submitted on 9 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
Predication Transformation: A Parallel Corpus-Based Study of Russian Verbless Sentences and Their English Translations

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 through history. In the present 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 ellipses, 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 ellipses (McShane, 2000). Nonetheless, verbless sentences are also found in English across all sentence types, including typical exclamatives (3) What a picture!, imperatives (4) And now to business., questions (5) What about my parental blessing?, and assertions (6) So much, then, for the introduction.

The paper opens with a description of the corpus. Part 3 outlines the methodology that involves automatic retrieval of verbless sentences. The definitions used for the automatic and manual annotation are discussed in Part 4. This section also provides a sketch of Hengeveld’s (1992) notion of verbal and non-verbal semantic predication, which we develop to include a distinction for antecedent-based ellipsis. In the final part, we present the results in terms of descriptive statistics, paying 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 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 stresses that the important role Dostoyevsky gives to dialogue makes this novel of particular interest for linguists studying speech acts (1982, p. 672). Targeting the key features of reliable parallel corpora described by Thomas Stolz (2007), we selected this work for 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 translations in future work.

The scope of the corpus is limited to the first fourteen chapters of the novel, 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 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 still remains a challenge. Studies of existing parsed corpora show that very often verbless sentence extraction is hindered 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 by developing an alternative method of automatic retrieval. Our approach involves customizing the automatic sentence segmentation process, 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 have achieved an average accuracy of 94% in terms of automatic recall of verbless sentences as compared with manually extracted results (Bondarenko, in press).

4. ANNOTATION

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 then separated into utterances and each verbless utterance and corresponding translation were manually annotated for the presence or
absence of a verb, the presence or absence of an antecedent-based verbal ellipsis following Mikhail Kopotev (2007a) and Marjorie McShane’s (2000) discussion of ellipsis types and licensing conditions, as well as verbal or non-verbal predication in accordance with Hengeveld’s (1992) model, modified for ellipsis as described below.

4.1. Hengeveld’s definitions and adjustment for ellipsis

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, p. 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, p. 26). Therefore, on Hengeveld’s conception, sentences containing a verbal, but semantically empty, copula are treated as instances of non-verbal predication.

In the same vein as Emile Benveniste (1966, p. 163) and Rodney Huddleston and Geoffrey Pullum (2002, p. 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 одушевясь, 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 a sluicing ellipsis in both languages and also 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.”]
“What dream?”

Since the ellipted antecedent is the semantically meaningful finite verb значит (znachit, 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.

We illustrate the distinctions between predication and sentence type in Figure 1, which is based on Hengeveld’s (1992, p. 27) diagram. Our version includes the modification for antecedent-based ellipsis, as well as examples from the present corpus.
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
&quot;verbalization rate&quot; concerns the change in the realization of the grammatical category of the verb, the &quot;predication verbalization rate&quot; 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) &quot;verbal predication&quot;, involving a non-copula verb (or a non-copula verbal antecedent), or (b) &quot;non-verbal predication&quot;, 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 &quot;be&quot; 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.

a) &quot;Idiomatic expressions&quot;

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 the meaning of the sentence transparent (Kopotev, 2015, p. 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 Фёдорович не ответил. − Инь ведь ты! − помолячив две минуты, проговорил опять Федор Павлович, косясь на сына: − сам ведь ты весь этот монастырь затеял, сам подстрекал, сам одобрял, чего ж теперь сердишься?] ['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 25.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 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 (for a discussion of the role of the verb in the translation of verbless questions see Bondarenko and Celle, in press).

b) 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 content of the Russian noun.

c) 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
CONJ 3S.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 –

сладострастник.
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 to the proper noun. The introduction of the verb, resulting in predication transformation, is a means of creating emphasis and keeping the information structure intact.

d) 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.
e) 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 кабак 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 Ракitin.]

[“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.

6. CONCLUSION

The corpus-based analysis of predication 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 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 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


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

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 jakikolwiek 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 predykacji w tłumaczeniu. Po automatycznej segmentacji, morfosyntaktycznej adnotacji i ekstrakcji zdań bezczasownikowych i ich tłumaczeniowe odpowiedniki są ręcznie przypisywane do typu predykacji werbalnej i niewerbalnej zgodnie z definicjami Hengevelda (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 predykacji nie jest stabilne międzyjęzykowo.

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