THE DISCURSIVE CONSTRUCTION *ДЕЛО В ТОМ, ЧТО* AND ITS PARALLELS IN OTHER LANGUAGES: A CONTRASTIVE CORPUS STUDY[1]

Dobrovol'skij D. (dm-dbrv@yandex.ru)

Russian Language Institute, Russian Academy of Sciences, Moscow, Russia

Pöppel L. (ludmila.poppel@slav.su.se)

Stockholm University, Department of Slavic and Baltic languages, Finnish, Dutch and German, Stockholm, Sweden

The primary goal of the present study is to improve methods for contrastive corpus investigations. Our data is the Russian construction *дело в том, что* and its parallels in English, German and Swedish. This construction, which appears to present no difficulty for translation into other languages, is in fact language-specific with respect to at least one parameter. It displays a large number of different parallels (translation equivalents) in other languages, and possesses a complex semantic structure. The configuration of semantic elements comprising the content plane of this construction is unique. The empirical data have been collected from the corpus query system Sketch Engine, subcorpus OPUS2 Russian, and the Russian National Corpus (RNC). The analysis shows that the construction *дело в том, что* has more than 50 parallels in English, over 30 in German, and about 30 in Swedish. In all three languages the most common means of translating the construction is to omit it. Also frequent are the English equivalents *the fact/thing/point/truth is (that)*; *(it's/this/that is) because*; the German expressions *nämlich*; *die Sache ist, die*; *denn*; and the Swedish constructions *saken är den att*; *problemet/faktum är att*. The semantic structure of *дело в том, что* includes the following components: 1) substantiation of something stated previously; 2) indication of the reason something has happened; 3) emphasis on the significance of what has been stated. The different translations of the construction are motivated by the fact that each specific context focuses on one of these meanings.

Keywords: parallel corpus, construction, contrastive corpus analysis, Russian, English, German, Swedish, language specificity, semantics

1.   Introduction

At first glance, the construction *дело в том, что* appears to present no difficulty for translation into other languages. It seems to possess complete compositionality, and even its literal translation is confirmed in the parallel corpora. Cf. the following examples from Russian-English (1), Russian-German (2) and Russian-Swedish parallel corpora.[2]

(1)   ***Дело в том, что*** *редактор заказал поэту для очередной книжки журнала большую антирелигиозную поэму.* [М. Булгаков. Мастер и Маргарита]

   The thing was that *the editor had—commissioned from the poet a long anti-religious poem for the next issue of his journal.*

(2)   *Ну, все равно, вы простите меня, — продолжал он, — но **дело в том, что** это ужасно, ужасно, ужасно!* [Л. Толстой. Крейцерова соната]

   *Nun, gleichviel, nehmen Sie es mir nicht übel,—fuhr er fort,—aber **die Sache ist** eben **die, daß** das alles so entsetzlich, entsetzlich, entsetzlich ist!*

(3)   ***Дело в том, что****... в последнее время, когда бы я ни пытался... что бы я ни сказал, все ее выводит из себя ... как будто она вообще не хочет слушать. Как будто она орет на меня постоянно. И я не знаю почему.* [Sketch Engine]

   Saken är den*… när jag väl pratar blir hon otålig, som om hon egentligen inte ville höra. Det är som om hon är förbannad på mig och jag vet inte varför.*

The English expression *the thing was that,* the German *die Sache ist die, daß* and the Swedish *saken är den* are practically word-for-word translations of the Russian *дело в том, что.*

2.   Research goals and data

The goal of the present study is to improve methods for contrastive corpus investigations. The following hypotheses will be tested on the basis of the materials of parallel corpora:
   (1) The Russian expression *дело в том, что* has a large number of various parallels in other languages, and the choice of each variant depends on specific contextual conditions.
   (2) Despite its apparent simplicity, *дело в том, что* has a complex semantic structure. The configuration of semantic elements comprising the content plane of this construction is unique.

---

[2]   Examples (1) and (2) are from the parallel corpora of the RNC, while example (3) comes from the parallel corpus OPUS2 Russian of the query system Sketch Engine.

(3) The expression *дело в том, что* is language-specific with respect to at least one of its parameters.

The empirical data have been collected from the corpus query system Sketch Engine, subcorpus of parallel texts OPUS2 Russian, which contains 307709872 tokens (15.02.2016); the Russian National Corpus (RNC), Russian—English, English—Russian, Russian—German and German Russian subcorpora of parallel texts.

3. Theory and methods

Analysis of the corpus data allows us to identify only the degree of variety in the means of translating a given expression into other languages. When one or another expression lacks a generally accepted standard context-independent translation equivalent, we can speak of an absence of systematic equivalents, i.e., a kind of non-equivalence. Whether such non-equivalence is connected with the category of language-specificity remains an open question.[3] Some interesting thoughts on this subject are presented in [Shmelev 2015], where the following parameters of language-specificity are identified. The first parameter is connected with the number of languages which lack the given phenomenon. The more such languages are discovered, the more possible it is to consider the phenomenon language-specific. The second parameter consists in the specificity of their content aspect (including connotations, background components of meaning, etc.), from which it follows that the degree of distinctiveness of the semantic configuration of an expression is directly proportional to its degree of language-specificity. The third parameter is immediately connected with the second: the more distinctive the semantic configuration of a lexical unit, the more difficult it is to find an adequate translation equivalent of this unit in another language. Here, as Shmelev notes, it must be kept in mind that the object of translation is not individual words but texts, so that the translator can deviate from exact equivalence on the lexical level. Such deviations do not necessarily imply that the translated units are language-specific. Nevertheless, it is natural to interpret the proposal of a large number of different translation equivalents as indicating the absence of a systematic equivalent. This allows us to measure quantitatively the degree of language-specificity in accordance with this third parameter, which is in fact at the focus of the present study.

The method of our analysis is based on considerations presented in [Buntman et al. 2014]. What it essentially entails is determining how many translation equivalents exist for potentially language-specific lexical units and then evaluating their dispersion. The expedience of calculating dispersion as a means of determining the statistical value of the scatter is in need of further substantiation. Other statistical instruments may prove to be more adequate.

The Russian expression *дело в том, что* was submitted to Sketch Engine, and we searched OPUS2 Russian for the construction and its translation equivalents

---

[3] On language-specificity see [Wierzbicka 1992, 1996; Zaliznyak, Levontina, Shmelev 2005, 2012; Zaliznyak 2015, Shmelev 2002, 2014, 2015].

in English, German and Swedish. Because there were very few parallel texts in all four languages, the search was done for each specific pair of languages: Russian and English, Russian and German, Russian and Swedish. We processed our findings manually to avoid information noise. The search system does not indicate from which or into which language a given context was translated. To ascertain the means by which the Russian construction *дело в том, что* is translated into other languages and the possibilities of this expression appearing in translation from other languages into Russian we used the parallel corpora of the Russian National Corpus (RNC)—English-Russian, Russian-English, German-Russian and Russian-German. Only the corpus query system Sketch Engine contains Swedish-Russian and Russian-Swedish corpora. Let us move on to our findings.

4.   Results and discussion

We begin the discussion of results with a presentation of the English results in Sketch Engine. Cf. Table 1.[4]

Table 1. *Дело в том, что*: English parallels Sketch Engine[5]

English parallels	Frequency
zero equivalence	154
the fact is (that)	123
the thing is (that)	98
the point is (that)	70
(it's/this/that is) because/because of	40
it's just (that)/it's that/just/this is that	27
in fact	26
the truth is (that)	26
however	16
the fact of the matter is (that)	15
indeed	13
the problem is (that)	12
you see	9
the reason is (that)	8
as a matter of fact	5
for	5
it's/this is about	5
it happens that/as it happened/what has happened is/what is happening is	5

---

[4]   Single hits are not shown in any of the tables.

[5]   The tables do not take into account differences in verb tenses. *The thing is (that)*, for example, includes contexts with all other tense forms.

English parallels	Frequency
the matter is (that)	4
but	4
since	4
it's a fact that	4
well	3
basically	3
what's true is (that)/it was true (that)	3
the consequence is (that)	3
the truth of the matter is (that)	2
the answer is (that)	2
the concern is (that)	2
the crux of the matter is (that)	2
the question is (that)	2
as	2
you know	2
look	2
the position is (that)	2
the thing about	2
in effect	2

Besides the data presented in the table, more than 43 single English correlates of the construction *дело в том, что* were found: *the situation is*; *that means that*; *my story is*; *the issue is*; *the reality is*; *the content is*; *the explanation is*; *the fact remained that*; *the fact that*; *this is due to*; *it has everything to do with*; *what I'm trying to say is that*; *except that*; *that is*; *in reality*; *actually*; *in practice*; *the word is*; *the plan was*; *here's the thing*; *this is the situation*; *sort of*; *the point being*; *the purpose of*; *it is not that*; *thus*; *it should be noted that*; *in truth*; *for the reason that*; *as it was*; *rather*; *in that it is*; *that is*; *instead*; *namely*; *in that connection*; *in this regard*; *it is which*; *to be blunt*; *here too*; *it is a matter of*; *accordingly*; *the trouble is*.

The first thing that stands out here is the large number of different English parallels of the Russian construction. In all we found 80 such equivalents.[6] The second important feature is that of these 80, 43 occur only once, which indicates significant scattering in these English parallels.

Let us now turn to the Russian-English parallel corpus of the RNC. Cf. Table 2.

---

[6] Some of them have been analyzed in [Aijmer 2007; Delahunty 2011, 2012].

Table 2. *Дело в том, что*: English parallels, RNC Russian—English

Engish parallels	Frequency
zero equivalence	27
the fact is (that)	14
the thing is (that)	14
the point is (that)	10
you see	3
actually	2
in point of fact	2
the matter is (that)	2

In this corpus we found 26 translation equivalents, of which 18 occur only once. It is evident from the table that the results partly coincide and partly diverge. Four of the most frequent equivalents—zero equivalent, *the fact is (that)*; *the thing is (that)* and *the point is (that)*—completely coincide, which indicates that the findings are non-random. At the same time, the relatively frequent constructions found in Sketch Engine—*in fact*; *the truth is (that)* and *however*—do not occur in the RNC, but *(it's/ this/that is) because/because of*; *it's just (that)/it's that/just/this is that* and *the fact of the matter is (that)*—occur only once. These divergences are quite natural. Sketch Engine is much larger than the RNC, while the RNC is much cleaner. In addition, in Sketch Engine it is impossible to determine which language is the source language, and the texts in these corpora differ with respect to genre. The RNC contains almost exclusively fictional texts, whereas non-fiction dominates in Sketch Engine.

Now for our analysis of the German materials. Cf. Tables 3 and 4.

Table 3. *Дело в том, что*: German parallels, Sketch Engine

German parallels	Frequency
zero equivalence	19
die Sache ist die (dass)	8
aber	5
es geht darum, dass	4
es ist (doch) so, dass	3
die Wahrheit ist, dass	3
wissen Sie	2
nur (dass)	2
Tatsache ist (nun mal)	2
es ist nur (dass)	2
ich meine	2
der Punkt ist	2

Table 4. *Дело в том, что*: German parallels, the RNC Russian—German

German parallels	Frequency
die Sache ist die (dass)	18
zero equivalence	11
nämlich	9
es handelt sich darum, dass	3
die Hauptsache ist, (dass)	3
doch	2
der Grund (war, dass)	2
es kommt (vielmehr/doch nur) darauf an	2

Although the Russian-German materials are considerably smaller in scope, the results exhibit tendencies similar to those observed in the English materials. In Sketch Engine we found 20 German parallels and in the RNC 13, some of which coincide and some do not. Two of the most frequent parallels in Sketch Engine—zero equivalent and *die Sache ist die (dass)*[7]—coincide with the most frequent ones in the RNC, although in reverse order. The most important difference is the absence of *nämlich* in Sketch Engine, whereas in the RNC it occurs 9 times. This difference is significant because even a superficial analysis of the word *nämlich* shows that its communicative function is very close to that of the Russian construction *дело в том, что*. On the whole, the German parallels in Tables 3 and 4 display considerable scatter. What most convincingly argues that *дело в том, что* is non-equivalent with respect to German is the partial absence of a translation equivalent in the parallel texts.

Table 5 shows the data from Sketch Engine, which is the only corpus of texts at our disposal containing Swedish parallels.

Table 5. *Дело в том, что*: Swedish parallels, Sketch Engine

Swedish parallels	Frequency
zero equivalence	45
saken är den att	16
men	8
problemet är att	7
faktum är att	4
det viktiga är (att)/det är viktigt att	4
det är för att	4
sanningen är att	3
grejen är den att	3
poängen är att	3
för (att)	3
det handlar om att	2

---

[7]   For analysis of *die Sache ist (dass)* cf. [Günthner 2008].

Swedish parallels	Frequency
det vad jag vill säga är att	2
i själva verket	2
jag/han menar att	2

We found 25 Swedish parallels, the most frequent of which—zero equivalent and *saken är den att*—once again testify to a certain non-translatability of the Russian expression *дело в том, что*. In the intermediate zone (from 2 to 13) there are 13 equivalents, 10 equivalents are used only once. Here as well we can speak of considerable scatter; that is, relative to Swedish the construction *дело в том, что* is difficult to translate.

Shmelev [2015] points out that for determining the language-specificity on the basis of the parameter of non-equivalence, the parallel corpora in which the language of the expression under analysis is the target rather than the source language have the best diagnostic potential. The appearance of such words is evidently the unconscious solution of the translator as a native speaker. For that reason we turned to the English-Russian and German-Russian parallel corpora of the RNC. Cf. Tables 6 and 7.

Table 6. *Дело в том, что*: English—Russian, the RNC

English parallels	Frequency
zero equivalence	38
the fact is (that)	36
for	34
it's just (that)/it's that/just/this is that	16
(that is) because	14
(as) you see	11
well	7
the thing is (that)	7
but	5
it happens (that)	4
actually	4
the truth is (that)	4
the point is (that)	4
in fact	4
the reason is (that)	3
the problem is (that)	2
I mean	2
as a matter of fact	2
I tell you	2
in truth	2
to begin with	2

The English-Russian corpus shows even greater scatter among the English equivalents of *дело в том, что*: in all, there are 54 different translations, of which 6 equivalents occur more than 10 times each, 15 are found in the range from 2 to 10 times, and 33 occur only once. It is reasonable to compare these findings with those from the Russian-English parallel corpus of the RNC (Table 2), since they are basically comparable. The corresponding results from this corpus are as follows: 3 are used more than 10 times each, 5 in the range from 2 to 10 times, and 18 occur only once. Of the most frequent equivalents, only two—zero equivalent and *the fact is (that)*—coincide. Our assumption that translators from Russian to English most often follow the form of the original, using constructions such as *fact is (that)*; *the thing is (that)* and *the point is (that)*, is fully corroborated. Translators from English to Russian, on the other hand, more often employ the language-specific expression whenever it is not dictated by form. Thus the most frequent group in Table 6 includes lexical units such as *for, just, because, you see*. In addition, syntactic means such as cleft sentences are also used. Cf. (4).

(4) *They didn't believe me at first either, he said. **It's just that** we get a lot of calls like this. But I believe you, Doctor, so why don't you continue with the story?*
[M. Connelly. City of Bones]

*Они тоже мне сначала не верили,—промолвил Гийо.—**Дело в том, что** мы получаем много подобных звонков. Но я верю вам, доктор, и, прошу, продолжайте.*

A comparison of the German-Russian and Russian-German parallel corpora of the RNC produces very similar findings. Cf. Table 7 and Table 4 above.

Table 7. *Дело в том, что*: German—Russian RNC

German parallels	Frequency
nämlich	27
zero equivalence	11
die Sache ist die, (dass)	10
denn	8
eben	3
aber	3
es kommt darauf an	2

The following features stand out. The formal correlate *die Sache ist die, (dass)* dominates in translations from Russian to German, while in the German-Russian corpus the word *nämlich* often correlates with *дело в том, что* although their formal structures have nothing in common. This confirms what was stated earlier. Cf. (5).

(5) *Wir landeten, und—die ganze Insel bestand aus einem großen Käse. Wir hätten dies vielleicht gar nicht entdeckt, wenn uns nicht ein sonderbarer Umstand auf die Spur geholfen hätte. Es war **nämlich** auf unserm Schiffe ein Matrose, der eine*

natürliche Antipathie gegen den Käse hatte. [G. A. Bürger. Die Abenteuer des Freiherrn von Münchhausen]

Как выяснилось, весь остров представлял собой большой сыр. Возможно, мы даже не заметили бы этого, если бы не одно обстоятельство, открывшее нам истину. **Дело в том, что** *у нас на корабле находился матрос, отроду испытывавший отвращение к сыру.*

Another feature of the German-Russian corpus is that the group of relatively frequent parallels includes the causal word *denn*, which is similar to the English conjunctions because and for in the English-Russian corpus.

5. Conclusion

The study advanced three hypotheses, each of which may be considered confirmed. The data presented in the investigation show that the construction *дело в том, что* has many different English, German and Swedish translation equivalents. It is important to note that most of these equivalents are not synonymous among themselves, which means that the choice of one or another of them depends not on the subjective preferences of the translator, but on contextual factors. Following from this as well is the fact that the construction *дело в том, что* has a complex semantic configuration that is unique relative to English, German and Swedish. The semantic structure of *дело в том, что* includes at least the following meanings: 1) substantiation of something stated previously; 2) indication of the reason something has happened; 3) emphasis on the special significance of the following clause. Equivalents from the various groups are selected depending on which meaning is being highlighted in the utterance. This can be illustrated with the following examples from Sketch Engine and the RNC.

1) English *you see* in **Дело в том, что** это ваша мать—**You see**, *it's your mother*), German *nämlich* in **Дело в том, что** она была значительно старше меня—*Sie war* **nämlich** *bedeutend älter als ich* or Swedish *det var så att* in *Det var så att jag råkade kalla henne subba*—**Дело в том, что** я случайно назвал ее кошелкой. Russian *дело в том, что* not only substantiates previous statements but also indicates the reason in each statement and emphasizes the significance of the following clause.

2) Causal conjunctions and constructions such as English (*it is*) *because, the reason is that*, German *denn*, and Swedish *för*. In the following examples English *it is because*, Swedish *för* and German *denn* indicate the reason for something that has been stated in the previous utterance: *То есть,* **дело в том, что** *я — женщина!*—*So* **it is because** *I'm a woman*; *Отличная книга. Её должен прочесть каждый. Видишь ли,* **дело в том, что** *если мы будем беспечны...*—*Gutes Buch. Jeder sollte es lesen.* **Denn**, *weißt du, es geht darum, wenn wir nicht aufpassen...*; *Может быть, вы поможете мне...* **дело в том, что** *мне очень нужно с ним встретиться. Du kanske kan hjälpa mig,* **för** *det är viktigt att jag får tag i honom.*

3) Focusing particles and constructions such as English *the point is, the thing is*, German *eben, die Sache ist die, (dass)* or Swedish *saken är den att, poängen är den att.* English *the point is that* in **Дело в том, что** этот противник никогда не отступает—**The point is that** this enemy never retreats; German *was die Hauptsache ist* in Но **дело в том, что** Анну я вам не отдам—*Aber* **was die Hauptsache ist**: ich lasse auf Anna nichts kommen; and Swedish *poängen är att* in Нет, но **дело в том, что** искушение всегда остается—*Nej, men* **poängen är att** frestelsen alltid är där focus on the importance of what is said. Russian *дело в том, что* does not only focus the following clause but conveys other meanings as well—substantiation and indication of the reason.

The uniqueness of the semantic and conceptual configuration determining the meaning of the construction testifies to its language-specificity relative to the languages examined here.

The findings of this and similar studies can be useful not only in developing the theory of language specificity but also in lexicography, translation and pedagogy. What we consider to be our primary accomplishment is that we have succeeded in outlining a new approach to working with parallel corpora in contrastive corpus investigations.

References

1. *Aijmer K.* (2007), The interface between discourse and grammar: *The fact is that*, in Connectives as discourse landmarks, Benjamins, Amsterdam, Philadelphia, pp. 31–46.
2. *Buntman N. V., Zaliznjak Anna A., Zatsman I. M., Kruzhkov M. G., Loshchilova E. Yu., Sitchinava D. V.* (2014), Informational technology in corpus-based studies: towards a cross-linguistic database, Informatics and its applications [Informacionnye tekhnologii korpusnykh issledovaniy: printsipy postroeniya krosslingvisticheskikh baz dannykh], Informatics and its applications [Informatika i ee primeneniya], vol. 8, issue 2, pp. 98–110.
3. *Delahunty G. P.* (2011), Contextually determined fixity and flexibility in *thing* sentence matrixes, Yearbook of Phraseology 2, Mouton de Gruyter, Berlin, N. Y., pp. 109–135.
4. *Delahunty G. P.* (2012), An analysis of *the thing is that S* sentences, Pragmatics, 22 (1), pp. 41–78.
5. *Dobrovol'skij D., Pöppel L.* (2015a), Corpus perspectives on Russian discursive units: semantics, pragmatics and contrastive analysis, Yearbook of corpus linguistics and pragmatics, Springer International Publishing, pp. 223–242.
6. *Dobrovol'skij D., Pöppel L.* (2015b), Russian constructions *то-то и N* and *в том-то и N* and their English and Swedish equivalents: a corpus-based cross-linguistic analysis, in Trends in Slavic Studies, URSS, Moscow, pp. 595–607.
7. *Dobrovol'skij D., Pöppel L.* (2015c), Entrenched lexical patterns: the Russian construction *в том-то и весь N,* Procedia—Social and Behavioral Sciences 206, Elsevier, pp. 18–23.

8.   *Günthner S.* (2008), "Die *die Sache/das Ding ist*-Konstruktion im gesprochenen Deutsch—eine interaktionale Perspektive auf Konstruktionen im Gebrauch, in Konstruktionsgrammatik II. Von der Konstruktion zur Grammatik, Stauffenburg, Tübingen, 157–177.

9.   *Shmelev A. D.* (2015), Russian language-specific lexical units in parallel corpora: prospects of investigation and "pitfalls" [Klyuchevye slova: perevod, parallel'nuy korpus, leksicheskaya edinica, semanticheskoe razlichie, lingvospecifichnost', "neperevodimost'"], Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2015" [Komp'yuternaya Lingvistika i Intellektual'nye Tekhnologii: Trudy Mezhdunarodnoy Konferentsii "Dialog 2015"], Moscow, pp. 584–595.

10. *Wierzbicka A.* (1992). Semantics, culture, and cognition. Universal human concepts in culture-specific configurations, Oxford Univ. Press, N. Y., Oxford.

11. *Wierzbicka A.* (1996). Semantics: primes and universals, Oxford Univ. Press, Oxford.

12. *Zaliznyak Anna A.* (2015), Russian language-specific words as an object of contrastive corpus analysis [Lingvospecifichnye edinicy russkogo jazyka v svete kontrastivnogo korpusnogo analiza], Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2015" [Komp'yuternaya Lingvistika i Intellektual'nye Tekhnologii: Trudy Mezhdunarodnoy Konferentsii "Dialog 2015"], Moscow, pp. 65–662.

13. *Zaliznyak Anna A., Levontina I. B., Shmelev A. D.* (2005), Key ideas of the Russian language picture of the world [Klyuchevye idei russkoy yazykovoy kartiny mira], Yazyki slavyanskoy kul'tury, Moscow.

14. *Zaliznyak Anna A., Levontina I. B., Shmelev A. D.* (2012), Constants and variables of the Russian language picture of the world [Konstanty i peremennye russkoy yazykovoy kartiny mira], Yazyki slavyanskoy kul'tury, Moscow.