

Solving Translation Problems with Aranea

Maria Kunilovskaya

University of Tyumen (Russia)

E-mail: m.a.kunilovskaya@utmn.ru

Abstract

This paper has a detailed description of a special course on the use of corpora in translation as part of BA programme in translation and translation studies. In the beginning we explain the ideology behind the course, and the way it fits the curriculum. The paper is structured around the practical tasks translators typically face at each stage of translation with regard to the translation direction (into or out of the mother tongue). We emphasize the practical skills that we aim at and provide multiple illustrations for the activities offered in this course. The paper is concluded with the analysis of the course results drawn from the students' final translation projects and their written reflections on the experience as well as from the post-course survey data.

Keywords: translator education; corpora in translation; translation solutions

1 Introduction

Since computer technology made various language databases available to the general user, those involved with language education are developing ways to introduce them to learners. The major goal of data-driven learning as a methodology - to develop "the ability of the learners to discover things for themselves" - formulated by Tim Johns (1993) in his groundbreaking work, is getting more appealing by the day as we witness an explosive growth of resources and tools available via the Internet.

In the early 2000s the benefits of corpus-based translator education were acknowledged in a number of publications, where the authors reported their classroom experiences and shared their approaches to using corpora for translation purposes (Aston 1999; Zanettin 2002; Zanettin, Bernardini, & Stewart 2003). These accounts as well as a multitude of later inquiries (most notably, Bernardini & Castagnoli 2008; Frankenberg-Garcia 2015; Pastor & Alcina, 2009) make it clear that - rephrasing a famous statement by Anthony Pym - corpora extend translator's capacities and corpus skills are not an option in today's professional world, but a necessity (Gil & Pym 2006). At the same time it is notoriously difficult to measure how effective discovery learning is (Boulton & Cobb 2016), while translators will indeed "tolerate the learning curve necessary to adopt corpora and concordancing software among their everyday working tools only if they derive benefits" (Zanettin 2002).

In this paper we offer a description of a special course on corpus skills designed to address typical translation problems. It aims to increase awareness of the existing language resources and equip translation students with basic corpora search skills. The course takes into account the results of the translation error analysis from the error-annotated subcorpus of Russian

Learner Translator Corpus,¹ which is also used to produce the course assignments.

The course is largely built around *Aranea*, a family of comparable gigaword web corpora (Benko 2014) and draws inspiration from the textbook by Thomas (2014). We rely on the students' translations and reflexive essays produced upon the completion of the course, and the data from a survey on corpora use in translation to generalize about the gains and limitations of the course.

2 Prerequisites and General Concept

Corpora in Translation Practice is offered to third year BA students majoring in translation and translation studies (English<>Russian) and involves 18 hours of contact time. Its primary goal is to demonstrate how available ready-made corpora can be useful for addressing a range of typical problems in general domain practical translation. Our previous teaching experience and results of this course show that the success of the course is contingent on the students' level of translation competence understood as "the underlying system of knowledge required to translate" in line with PACTE research (Beeby et al. 2011). It means that if we want to see effective application of corpus skills for solving translation problems, the students should be able to recognize the latter in the first place. This stance determines the place of the course in the curriculum and the activities offered.

It makes the second part of *Discourse Analysis for Translators*, a broader discipline, which aims to develop students' ability to understand the source text (ST). Students learn to interpret the ST as a whole, define functions and communicative relevance of its elements and perform contrastive analysis at different language levels. As part of the course ideology we insist that it is crucial (i) to have a coherent interpretation of the source and (ii) to envision functional as well as linguistic parameters of the intended target before producing any translational output. This arrangement in the curriculum is a conscious effort to develop a functional approach to translation and reduce the number of mechanic and literal renditions.

In developing the tasks, we proceed from three major assumptions about the translation process:

- it is a creative activity, not unlike engineering, aimed at finding most effective practical solutions that conform to a number of desired parameters; the ability to set the task requires critical thinking and analytical skills as well as linguistic and world knowledge;
- due to the underlying variety of language expression and interpretative potential of the source text, translation solutions are highly variable by nature; there can be no binary (right or wrong) assessment for them, but they can be described in terms of more or less acceptable in the given context (Pym, 1992); the variety of expression is closely matched by the number of ways to produce them;
- in many cases, the most creative part of the translation process is finding the most standard conventional way of conveying ideas (Riabtseva, 2014).

Corpus skills are not viewed as an end in itself in this course: they are instrumental for solving translation problems. We argue that they increase translator's awareness of the disparities between working languages and boost creativity by exposing translators to a range of real life

¹ www.rus-ltc.org/about.html

contexts that express ideas, similar to those in the ST. The course introduces students to a range of other resources, except *Aranea*. We begin with simple interfaces like *Just the word*² and *SkELL*,³ which help us to provide the minimum understanding of the corpus terminology. There is no theoretical component in the course, and the necessary corpus linguistics notions are introduced by and by, inside the respective tasks. After the first introductory module, we offer a formative test, which brings together the new concepts related to the corpora properties, query interfaces and neo-Firthian linguistics. These include *lemma*, *collocation*, *PoS tags*, *node*, *concordance*, *relative frequency*, *word sketch*, *wildcards*, *discourse prosody*, etc.

We demonstrate the importance of asking questions about meanings conveyed (ex. what is the difference between *pervasive practices* and *widespread practices*?) and emphasize the benefits of querying thousands of authentic contexts to enhance language competence. After all, when it comes to translating into a foreign language, you can never be too sure about anything. The students are tasked with investigating surprising instances of usage and cross-linguistic contrasts. For example, the discovery that in mass-media texts *to argue* is primarily used in contexts incompatible with its straightforward students'-top-of-the-head interpretation "express diverging or opposite views, typically in a heated or angry way" produces a steady wow-effect.

Generally, the course offers a series of test-like tasks to solve linguistic problems that were unsuccessfully tackled by the previous student generations. It exists as a Moodle class and is suitable for self-study. The tests are designed as formative learning activities, rather than for assessment. Some of them include open-ended tasks that require reflection rather than a definitive solution.

The course finishes with a hands-on overview of other available resources for English and Russian, including Mark Davies's *Corpus of Contemporary American English*⁴, the *Russian National Corpus (RNC)*⁵, *IntelliText*⁶, a translation project and a post-course survey.

In the rest of the paper we describe the syllabus and exemplify the activities offered in each module.

3 Teaching Basic Corpus Search Skills and Addressing Key Lexical Issues

The course opens with the demonstration of simple corpus interfaces like *Just the Word* and *SkELL*, followed by simple queries in *Aranea* via *NoSketch Engine* interface. Their functionality can be effectively used for exploring combinability and finding semantically similar words, and it is *combinability and lexical choice* that are the major issues in learner translations (in both direction) as can be seen from the pie-charts in Figure 2. The assignments in this part of the course are built with two objectives in mind: (i) they introduce the resources and (ii) typify the standard linguistic inquiries that translators make. Naturally,

² www.just-the-word.com/

³ ruskell.sketchengine.co.uk/run.cgi/skell

⁴ corpus.byu.edu/coca/

⁵ www.ruscorpora.ru/

⁶ corpus.leeds.ac.uk/itweb/htdocs/Query.html#

the tasks need to be solvable with the resources at hand at least at the initial stage.

Examples 1-4 demonstrate assignments from this part of the course. They contain less adequate learner translations after the asterisk and an improved corpus-informed version in the square brackets; glosses per se are offered in the double quotes, where necessary.

- (1) (from a political speech) *Мы приветствовали такое начинание* - *we greeted this initiative [-> we welcomed this idea]

Students are asked to look through verbs that take *initiative* (the candidate head word of the phrase) as an object and suggest a more idiomatic replacement for *greeted*. Then we try similar words to find possible alternatives for the head word. Other examples for working on collocations in both directions include *пугливая рыба* - *fearful fish [shy fish], *to break monopoly* - *сломить монополию [уничтожить монополию], *to curb the misuse of the data* - *обуздать злоупотребление [прекратить/предотвратить злоупотребление данными]. To illustrate possible search settings for building an informative concordance, we limit the initial search for “fish” as the node by the contexts including all variants for *пугливый* from Multitran (a bilingual dictionary, popular with the students) “timid skittish shy fearful scary” in the left window of size 1.

The screenshot shows the NoSketch Engine search interface. The top navigation bar includes 'Home', 'Search', 'Word list', 'Corpus info', 'My jobs', and 'User guide'. The main search area is titled 'Araneum Anglicum II Maximum (Global English, 17.06) 11.4 G'. The search settings are as follows:

- Corpus:** Araneum Anglicum II Maximum (Global English, 17.06) 11.4 G
- Simple query:** [empty field] [Make Concordance]
- Query type:** simple (selected), lemma, phrase, word, character, CQL
- Lemma:** fish [PoS: unspecified]
- Phrase:** [empty field]
- Word form:** [empty field] [PoS: unspecified] [match case]
- Character:** [empty field]
- CQL:** [empty field] [Default attribute: word]
- Context:**
 - Lemma filter:** Window: left, 1 tokens. Lemma(s): timid skittish shy fea [any] of these items.
 - PoS filter:** Window: both, 5 tokens. PoS: Dt, Nn, Aj, Pn (Aj is selected).

Buttons for 'Make Concordance' and 'Clear All' are visible at the bottom of the search area.

Figure 1: Search settings to test combinability of the head word and modifiers from a bilingual dictionary

A possible approach to find appropriate translations for *сильно укусить* - *strongly bite requires arriving at an image of the desired result first. From what we know about the structure of English, we can expect a lexico-grammatical pattern *a delexical verb (have, take, make, give) + Adj + a deverbal noun (a bite)*. The query to return corpus examples for this pattern is built with CQL: [lemma="give"] [] [atag="Aj"] [lemma="bite"], and we are looking for “animal” contexts to identify the adjective typically used in them.

To expose the problem of collocation and at the same time to demonstrate the limitations of smaller corpora when it comes to lexical queries, the students are asked to back-translate and interpret a translation like (2). The source text includes the phrase *витиеватый туннель*

(“not straight tunnel”) which triggered the misleading italicized *flowery tunnel* rendition. It is clear that this translation variant results from unprofessional dictionary use and corresponds to a more typical use of the Russian adjective in collocations like *витиеватое выражение, витиеватый стиль* (“flowery expression, style”).

(2) (from a translation of the Sochi Aquarium advertising) [...] there is a hall that displays the inhabitants of the seas and oceans. It includes 13 small tanks and a huge 3 mln litres aquarium with *a *flowery tunnel* underneath.

The list of collocates for *tunnel* as a node built from concordances limited to *Aj + tunnel* pattern in *Araneum Anglicum Maximum* does include more adequate solutions, which might not occur to a neophyte translator (e.g. [...] those hundreds of kilometers of *twisted tunnels* and passages and chambers [...], [...] somewhere in the *twisted tunnels* that navigated the Roman catacombs [...], Crawl and climb through a maze of twisted tunnels and shafts to feel for yourself [...]).

We emphasize that search results are to be interpreted with regard to the parameters of the corpus behind the interface (size, web-data effects, temporal aspects of the sample, automatic processing errors) and the limitations of the interface itself (frequency cut-offs). For example, *engaging activity* is not registered in the *BNC* and has only two hits in *COCA*, while *Google Books Ngram Viewer*⁷ shows an explosive growth for this phrase since the late 1980s.

It is also worth remembering that some ideas expressed by a phrase in the source language can well be represented differently in the other. If the analysis of corpora hints in this direction, it makes sense to look for words similar to the head word (e.g. *мощная инфраструктура* (“developed infrastructure”) (about a country sports centre) > change to *facilities?* *влиться в коллектив преподавателей и студентов* (“to join the team of professors and students”) > change to *university community?*). It is unlikely that a student will make it from *хотя и не входит в должностные обязанности* (“it is not included into one’s official duties”) to *to perform above one’s pay grade* by querying corpora, but it helps to generate alternative hypotheses and compare them. And that is the core translation competence according to Рум (2003).

In some cases it turns out that a part of the phrase is excessive in the target language and the concept is conventionally conveyed by just the head word (e.g. *сборная команда области* (“joint team of the region”), *the world’s gainfully employed* - **оплачиваемых работников* (“paid workers”), **работающих за заработную плату* (“those working for pay”).

In terms of search skills by the end of this module, students realize that there are multiple ways to produce a list of the left immediate adjectival collocates of *growth* to explore possible renditions for *бурный рост* (“rapid growth”). They can find verbs that take *criticism* as an object to re-express *вызвать критику, подвергнуться критике* (“to attract criticism”) and explain semantic preferences of *to focus* as a transitive verb in active voice while excluding the most obvious and numerous collocation with *attention* to check whether a review can focus certain aspects of a movie, or does it bring them into focus?

⁷ <https://books.google.com/ngrams>

4 Corpora Solutions at Different Stages and Directions of Translation

4.1 Aranea Meets Error-tagged RusLTC: in and out

The central component of the course is focused on the *NoSketch Engine* interface and the *Aranea* web corpora. This combination seems so effective for our purposes that over the last few years we have redrafted the course to give *Aranea* priority and reduced the course components on the other corpus resources to an overview. *NoSketch Engine* offers a relatively user-friendly, but powerful and flexible interface, which allows most sophisticated searches, while the easily available gigaword *Aranea* provide enough data even for moderately specialized translation projects.

At the top-level the assignments in this module are arranged in 3 categories, following the translation process-centred (rather than search technology-centred) approach:

- solving translation problems when translating into a foreign language;
- addressing error-prone areas in translation into the mother tongue;
- using Aranea for contrastive analysis.

After discussing translation issues at hand and facilitating linguistic questions that need to be answered, we demonstrate more advanced NoSketch Engine functionality such as limiting concordances by PoS and lexical context, sorting and filtering initial results, building frequency lists for positional elements and recovering longer patterns through CQL.

Corpora in Translation Practice is an attempt to address real translation problems of the current student population by feeding the results of the prior translation quality assessment back into the training process. Over the last 5 years we have acquired a corpus of annotated learner translations, which yields statistics on 30 types of translation errors as well as examples for each type. The corpus is bi-directional. At the time of writing, its EN>RU part includes 599 texts (247,230 tokens), while RU>EN part has 166 texts (73,239 tokens).

The distribution of error types (see Figure 2) shows that major problems in English-to-Russian translation have to do with:

- lexical choice (including lexical combinability, 30.7%);
- ST comprehension (inability to reproduce reference, 23.7%).

In the opposite direction of translation most error tags fall into the categories of:

- lexical choice (32.6%);
- syntactic errors (including word-order, incomplete or ungrammatical structures and propositions, 22.4%);
- morphological errors (mostly determiners or lack of thereof, 12%) (almost on par with reference errors, 12.6%).

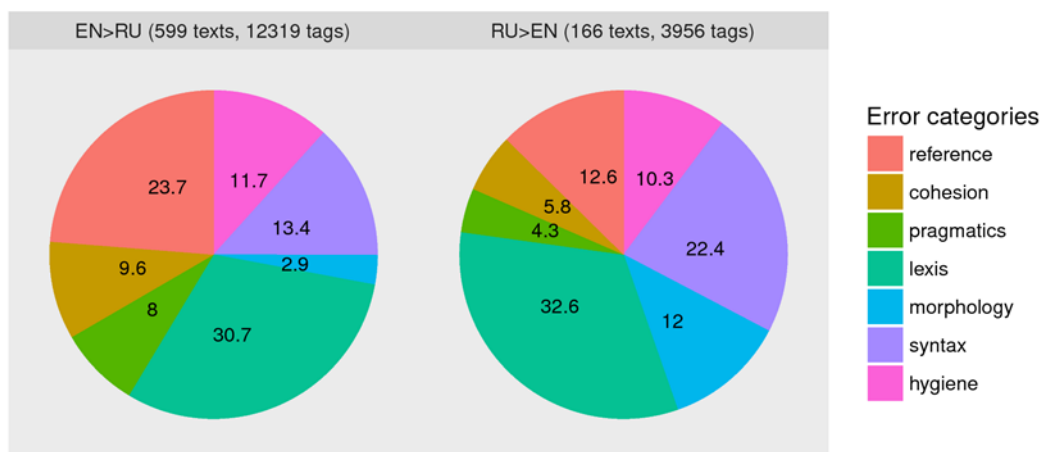


Figure 2: Distribution of major error categories in EN<>RU learner translations

Error statistics and examples from past translations help to further arrange tasks according to stages in translation process: ST comprehension, generating translation hypotheses and selecting the most adequate translation solution (editing). The first aspect of the three is more important in English to Russian translation, while in the opposite direction more attention is paid to grammaticality issues and active language skills.

4.2 Source Text Comprehension

It is intuitively reasonable that in translations from the students' L2 (English) into their mother tongue, more problems are associated with ST comprehension, including inability to re-express the same ideas (reference) coherently, to render intended text functions and implications (pragmatics) and ensure that the text "hangs together" in a meaningful way (cohesion).

Some problems at this stage can be addressed by exploring pragmatic prosody of the items, overtones of contextual meanings and by identifying the set character of phrases. These tasks usually require more sophisticated search skills and access to gigaword corpora. Example (3) is difficult to interpret, especially given that students have problems finding the phrase in italics in their favourite lexicographic resources:

(3) The *wide currency* of this philosophy of the "survival of the fittest" enables people who act ruthlessly and selfishly [...] to assuage their consciences by telling themselves that they are only obeying a law of nature.

Araneum Anglicum Maxicum has 273 contexts for the phrase, which help to interpret it as meaning something like "gain popularity, become known and accepted".

Student translations are notorious for literal rendition of set phrases. A typical example of "it is a good horse that never stumbles", which, if literally translated, can hardly be communicative in Russian, is commented by Komissarov, a prominent Russian translation studies scholar (Komissarov 1990: 64). Example (4) has an outright erroneous translation of *and a good thing, too* in a heading from *The Economist* article "Leviathan's Spyglass" (July 2010) and reveals lack of professional competence, which, perhaps, cannot be remedied by corpus skills. But we argue that alarming students to pratfalls like that is useful.

(4) The traditional census is dying, *and a good thing, too*. - *Традиционная

перепись умирает и все хорошее тоже (“traditional census is dying and everything good is dying, too”).

The analysis of the 51 extendible contexts for the phrase reveals that this item has a strong evaluative connotation, which is also positive. This example is convenient to show that punctuation needs to be separated by a comma in the NoSketch Engine phrase search field.

An effective method to demonstrate pragmatic limitations of some word-choices is to ask students to interpret translations without first looking into the source texts. Failure to do so without referring to the source or major deviations from the source signal inadequacy in translation, which can be related to limited understanding of the source, including on the lexical level. In example (5) the translator seems to have assumed that the English *border* is as polysemic as the corresponding Russian *граница*, which is likely to be false, judging from the analysis of the concordances. The English source is more about state borders and transnational nature of the tech monopolies like Google rather than generally about the alleged “limitless world” as their context.

(5) [...] the new internet giants *operate in a largely borderless world* where their main source of profit is intangible intellectual property. - *новые интернет-гиганты *действуют в безграничном мире*, где главным источником прибыли является интеллектуальная собственность (“new Internet giants act in a limitless world [...].”)

It can be argued that in the examples above translators lacked logic and textual competence, and querying corpora can hardly help with that. It is true that corpora are of little use when naive translators do not notice disparities between language systems, ignore cultural differences and, therefore, experience little difficulties while translating. There is no doubt that the more professional the translator is, the more they benefit from corpora, and that little learning is a dangerous thing. Our point is that exposing learner translators to corpora facilitates linguistic inquiry and encourages analysis, re-phrasing, thinking of patterns, alternatives and explanations, all of which is beneficial for the subsequent re-casting into another linguistic form.

4.3 Hypotheses Generation and a Quest for Idiomatic and Grammatical Renditions

The most popular part of the course is the one that deals with hypotheses generation in translation into L2 (English). On the one hand, students find it difficult to ignore problems in this translation setting, and on the other hand, this is when the creative nature of translation uses the discovery potential of corpus data analysis to the full. Corpora expose students to new patterns and broaden their repertoire of expression and flexibility. We entertain the idea that corpora can to an extent remedy the lack of systemic contrastive knowledge or, ideally, consolidate it. For example, searching for formal correspondences to the key components in *регулярно проводят конференции* (“regularly hold conferences”) can bring an inquisitive user to a more acceptable lexico-grammatical transform: *to hold regular conferences*.

Looking into discursive prosody of *to cause* through generalizing about emotional and evaluative connotations of its most frequent collocates, students realize that *but this topic did*

not cause much curiosity is not the best way to put it, which also explains the conspicuous absence of the combination in the biggest databases.

The error analysis indicates that translations into a foreign language have even more lexical issues and feature lower grammaticality. Morphology and syntax errors jointly are at 34.4% of all error types in RU>EN compared to 16.3% in translations into the mother tongue (EN>RU, see Figure 2). In this section we discuss several examples of discovery procedures that can help to arrive at appropriate lexical solutions when bilingual dictionaries are either inadequate or helpless, and comment on using corpora for grammaticality testing.

In solving lexical problems, the task is to focus contexts similar to the source text. The piece in example (6) is taken from a scholarly paper on the history of reading and geographical imagination.

(6) Поскольку текстовых маргиналий в моем распоряжении практически не было, особое значение имели *знаки на полях*. Само распределение и концентрация этих знаков позволяют представить себе карту читательского интереса к определенным темам. - *As there were practically no *text marginals* at my disposal, the *signs on the margins* were of particular importance. The distribution itself and *concentration* of these *signs* make it possible to *imagine a map* of reader's attention to some topics.

This example can be improved on several counts: i) there is an obvious paronymic mix-up of *marginals* and *marginalia*; ii) the choice of *signs* as a reference to one type of marginalia is more than doubtful; iii) the literal *concentration* rendition; and iv) a calque with *imagine*.

To narrow down the search to similar contexts in *Araneum Anglicum Maximum* we looked for occurrences of *reader* in the vicinity of *margin*. Reading concordances does not only help to come up with several possible renditions but also reveals differences between them, which is important in this case. *Marginalia* – is a general term which is used to refer to all types of readers' annotations in the margins, *reader's notes* are more likely to be used to refer to verbal reactions, while *marks* describe symbols like stars, ticks and underlining. At the same time the word “sign” is not used in these contexts. Looking through concordances helps to pick authentic collocations in this domain that can fit in other parts of the same text.

Several student translators of this passage do not move on from the literal *concentration* variant. A quick look-up in *SkELL* similar words and cross-checking possible alternatives (number, amount, density) with *Aranea* for better fit to the context of reader's marks can do the job in no time.

The last improvement requires additional search for *map* (unspecified PoS) in the context of *interest*, which shows that *imagine* and *to some topics* are probably excessive, and we can go for “the distribution and density of the marks help to map the readers' interest and [...]”.

Sometimes dictionaries fail to list a usable translation candidate. If we want to describe *books in matching covers* (books that come from one and the same series; the source: *книги в одинаковых переплетах*) and look up the adjective, we get only *equal*, *the same*, *identical*, *alike*, *uniform*, but a gigaword corpus does not fail to return *matching* in contexts that have both *book* and *cover* as search anchors.

More creativity is required to find an authentic context similar to *попал в жернова*

гражданской войны и больше не вернулся (“got into the millstones of the Civil war and never came back”). The obvious plain and neutral variant is *disappear during the war*. However, if we want a more figurative and expressive candidate, we can try a synonym for the verb (e.g. *vanish* from “vanish into thin air”) in the vicinity of *war*. As a result, we get “vanished in the fury of the Civil war” instead of the students’ arguably less effective “was consumed by the brunt of the civil war” and “met the events of Civil war. He did not come back”.

In a similar manner one can find the name for enclosed areas where big animals are kept in zoos (*вольер для крупных животных*, not cages!). Searching for the pattern “N + and history of the region” will return a better interpretation for the first noun in the phrase [...] *дает представление о природе и истории края* (“[...] gives a general impression of the nature and history of the region”).

Turning to grammaticality issues, we stress the imperative to doubt and double-check on “little” words like prepositions and determiners as well as on syntactic restriction that some words have. In many cases several variants are possible or even equally frequent and the choice is made on the basis of semantic properties of the contexts (e.g. *at/on (a) par with? smb's interest in/to smth, when compared to/with/against smth, to be the/an end in itself*).

At the end of the module we offer a test-like formative assignment, which overviews and consolidates the search skills learnt in the process of solving translation problems. For example, there is an activity which summarizes the meaning of regular expressions used in CQL queries: Match the special characters (dot, pipe, asterisk) in the query [lemma="recommend"] [lemma="to|for"] [tag="N.*|PP"] with their meaning. Analyse the results of the query. What observations can be made on the use of prepositions with *recommend*?

4.4 Contrastive Analysis and Frequency Calques

We emphasize the idea that a functional translation is supposed to approximate originally authored texts in the target genre. Therefore, comparable corpora are more useful than parallel corpora (such as *Linguee*⁸ or *Reverso Context*⁹) or bilingual dictionaries (such as *Multitran*¹⁰) for identifying idiomatic ways of expression and assessing alternative solutions. These resources come in handy on the previous stage of translation. In fact, one of our goals is to demonstrate which resources are more effective for finding which linguistic information. We encourage the use of dictionaries, parallel corpora and synonym finders for hypothesis generation.

Aranea project contains comparable corpora, which makes it a valuable resource for detecting and describing frequency calques – forms that are possible in the target language but not as frequent or idiomatic in target language non-translation. The usual suspects for English-Russian frequency calques include: *citizen/гражданин*, *worker/рабочий*, *territory/территория*, *юридическое лицо/juridical entity*, *обеспечивать/ensure*,

⁸ www.linguee.com/

⁹ context.reverso.net

¹⁰ www.multitran.ru/

inhabitant/житель.

We motivate students to have two comparable corpora open and compare relative frequencies in each word pair.

5 Course Results and Evaluation

The course effectiveness is assessed in two ways. We provide analysis of the students' translation practices explicitly reflected in a written report submitted at the end of the course and implicitly manifested in their independent translations within the final task. Secondly, we aggregate the results of the post-course survey that is meant to reveal what the students have learnt throughout the course. Over the last two years the course was taken by 23 students, and we managed to obtain complete feedback from 17.

The final project requirements include a practical translation of an informational text (EN>RU, 350-400 words) and a reflexive essay (translator's commentary). It discusses translation problems detected in the task and describes ways to find solutions. The students are also asked to offer and compare available translation candidates and put forward their motivation behind the final choice. The essays demonstrate that the students incorporated corpora into their translation routines, but they use them mostly for cross-checking variants, suggested by dictionaries or parallel databases. There are only few examples of applying corpora to the source text analysis or to generating translation hypotheses. If we focus on the problem areas rather than gains, there are two major considerations. First, students fail to adequately interpret data extracted from corpora and produce applicable knowledge: they can consider a form a valid translation on the basis of its presence in the corpora. In example (7) the plural form of the dependent noun is transferred from the source language into the target, though it is less common in Russian:

(7) has sparked *conversations* around the world - *стало поводом для *обсуждений*

In example (8) the student made a point of finding a single verb that can take three different objects (job transitions, sick relatives, new babies) and justifies the solution by pointing out the frequencies for each in the *RNC*. But the final collocation has connotations that make it less fit for this context. It is used in contexts where adults make efforts to organize kids' life or teach them behave, and is less typical for reflecting financial aspects of having children.

(8) This would provide a basic income to all Switzerland's residents, allowing them to survive through job transitions, caring for a sick relative, new babies, and throughout their lives. - *Таким образом, каждый житель Швейцарии получал бы гарантированный доход, который позволил бы успешно *справиться с* увольнением, болезнью родственника, *новорожденными детьми* - так на протяжении всей жизни.

Second, the test translations confirm that corpus skills are useless if translators do not see translation problems in the first place. Example (9) has a typical lexical calque.

(9) A growing popularity among *thought leaders* [...] - *Растущая популярность среди *лидеров мысли* [...]

A more idiomatic reference to a recognized authority in a field, especially when it comes to social and political issues, is “лидеры (общественного) мнения” in today's Russian mass

media. The offered collocation seems to exist only with the adjectival specification of the field of expertise: лидер гуманистической / дизайнерской / философской мысли.

The post-course survey returned moderately optimistic results. Just one student admitted a failure on the course due to the steep learning curve. Over 80% of students added corpus resources to their professional bookmarks. About 50% claim to have registered for gigaword *Aranea* and would use this resource to explore usage in the English language. At the same time when it comes to inquiries into their mother tongue, they give preference to the national corpus. Bilingual dictionaries rank higher in terms of popularity as a source of linguistic information, but corpora outperform monolingual dictionaries, and surprisingly *Aranea* does better than *SkELL*. As for the search skills, on the final test 71% of students were able to employ simple pattern search, 59% managed to produce required collocate lists, while 53% knew how to filter and sort concordances. The overall usefulness of the course is estimated at 80% on average, with the practical skills focus, lively in-class experience and test-based learning being pointed as the major strengths. The top three complaints include lack of feedback, (surprisingly) lack of theory and annoying errors in the tests.

6 Conclusion

Corpora in Translation Practice is designed within the constructionist paradigm in education: though it uses direct instruction as a teaching method (at the early stages), it aims at developing independence and has open-ended inquiry tasks, which motivate students to construct their own ways to solve translation tasks. It is based on solving real life problems, which encourages heuristic approaches and critical thinking; it is open to multiple solutions and integrates corpus skills onto translators' professional competence.

We are fully aware that mastery of corpus skills does not guarantee acceptable translation – there are other important aspects to translation competence that can hardly be tackled via corpora (coherence issues, implicatures, intertextuality), but we argue that translation (esp. into a foreign language), is a strong case for the use of corpora. They encourage linguistic inquiry and provide grounds for motivated and reliable language choices as well as prevent awkward solutions at least at the lexical and morphosyntactic levels.

7 References

- Aston, G. (1999). Corpus use and learning to translate. In *Textus Online Only*, 12(2), 1000--1025.
- Beeby, A., Fernández, M., Fox, O., Albir, A. H., Neunzig, W., Rodríguez, P., & Romero, L. (2011). Results of the validation of the pacte translation competence model: Translation project and dynamic translation index. In *Cognitive Explorations of Translation*, 30–53.
- Benko, V. (2014). Yet Another Family of (Comparable) Web Corpora. In P. Sojka, A. Horak, & K. Pala (Eds.), *Text, Speech and Dialogue. 17th Internat. Conf., TSD 2014* (pp. 257–264). Springer International Publishing.
- Bernardini, S., & Castagnoli, S. (2008). Corpora for translator education and translation practice. In *Topics in Language Resources for Translation and Localisation*, 39–55.
- Boulton, A., & Cobb, T. (2016). A meta-analysis of DDL research 1: Rationale, methodology and outcomes. In Proceedings of the 12th Teaching and Language Corpora Conference (TaLC), Giessen, Germany: Justus Liebig University.

- Frankenberg-Garcia, A. (2015). Training translators to use corpora hands-on: Challenges and reactions by a group of thirteen students at a UK university. In *Corpora*, 10(3), 351–380. <http://doi.org/10.3366/cor.2015.0081>
- Gil, J. R. B., & Pym, A. (2006). Technology and translation. In *Translation Technology and Its Teaching*, 5–19. <http://doi.org/10.1002/9781405198431.wbeal1222>
- Johns, T. (1993). Data-driven learning: An update. In *TELL & CALL*, 3, 23–32.
- Komissarov, V. N. (1990). Theory of translation (linguistic aspects) [Teoriya perevoda (lingvisticheskiye aspekti)]: A university textbook. Moscow.
- Pastor, V., & Alcina, A. (2009). Search Techniques in Corpora for the Training of Translators. In *Proceedings of the Workshop on Natural Language Processing Methods and Corpora in Translation, Lexicography, and Language Learning*, 13–20.
- Pym, A. (1992). Translation error analysis and the interface with language teaching. In *The Teaching of Translation*, 279–288.
- Pym, A. (2003). Redefining Translation Competence in an Electronic Age. In *Defence of a Minimalist Approach*. In *Meta: Journal Des Traducteurs*, 48(4), 481. <http://doi.org/10.7202/008533ar>
- Riabtseva, N. K. (2014). *Applied Translation Studies. Linguistic aspects*. Moscow: Flinta.
- Thomas, J. (2014). *Discovering English with the Sketch Engine*. Brno: Laptop Languages.
- Zanettin, F. (2002). Corpora in Translation Practice. In *Language Resources for Translation Work and Research LREC 2002 Workshop Proceedings*, 1, 10–14.
- Zanettin, F., Bernardini, S., & Stewart, D. (Eds.). (2003). *Corpora in Translator Education*. St. Jerome Publishing.

Acknowledgements

This work has been supported by the Russian Foundation for Basic Research within Project No. 17-06-00107/18.