Corpora-based Teaching of Specialized Italian-Russian Translation

Ksenia Balakina

University of Bologna
E-mail: ksenia.balakina2@unibo.it

Abstract

The paper stresses the important benefits that produce classroom concordancing activities and data-driven learning techniques applied in the classes of specialized translation into a non-native language. Under this study, the author concentrates on the particular way words are combined within the domain of commercial Russian, as in the case of inverse translation, the problem of words' combinability presents particular difficulties for Italian-speaking students. The paper stresses the importance of providing the students with an efficacious tool and method for autonomous search for accurate collocations of the non-native target language. For this purpose, the large web-based Aranea corpora were introduced in class and have proven to be an indispensable tool for the training of more independent and confident translators. The paper analyses some of the students' mis-collocations from a small learner corpus of translations and suggests some Aranea-based activities mirrored at collocations research and words' behaviour exploration. The exceptionally positive feedback received from students confirms the effectiveness of the approach.

Keywords: specialized translation into non-native language, translation teaching, collocations, classroom concordancing, web-corpora

1   Background

The present study was carried out within the Master's Degree course of Specialized Translation from Italian into Russian at the University of Bologna. During the course, the students are introduced to the strategies and techniques used in specialized translation and learn to apply them to the translation of technical and scientific texts of different genres and text types.

A significant part of the course was dedicated to the translation of commercial correspondence and documents, such as commercial proposals and contracts. The choice of the genre and domain was made in order to bring the students closer to the real world of work where the skills related to the translation of commercial documentation is frequently requested.

All the students of the class are Italian-speaking and their level of proficiency in Russian can be valuated as B2+/C1 according to the Common European Framework of Reference for Languages. On the basis of the students' translations, a small learner parallel corpus was compiled. All students have signed a consent form, giving permission for the use of their translations for research purposes.

2   Directionality

The course presents particular difficulties for Italian-speaking students, because they have to deal with the issues related to translation into a non-native language, namely, into Russian.

Despite the fact that training students to perform (inverse) translation into a non-native language (NNL) is often considered to be just a part of their academic studies, the survey conducted in 2014 by the International Association of Translators and Interpreters has confirmed that working in a NNL is a widespread practice among professional translators (IAPTI 2015). Under the survey, more than 50% of 780 translators from 80 countries declared to translate into a foreign-language on regular basis.

Translating into a NNL is associated with a higher level of difficulty for the translators that have to possess an advanced language competence. However, it is not the only factor that influences the quality of the translator's production. It is translator "confidence" that becomes one of the central issues to be taken into consideration by translation teachers (Kiraly 2000).

From a pedagogic perspective, it is also important to identify the areas that are more error-prone when translation into a foreign language is performed. The survey (IAPTI 2015) sheds light on this question as well, revealing that errors related to wrong collocation and usage of words are much more frequent among non-native speakers, compared to native speakers of the target language. The survey data confirms that the problem of words' combinability presents even more difficulties than the well-known issues related to grammar and terminology.

The lack of language intuition in the target language makes the choice between alternative language units much more difficult for a non-native translator, often resulting in a production characterized by some atypical lexical combinations that render the translation in L2 unnatural for a native speaker. Thus, collocation use is most likely to be one of the main features that makes the difference between translations in L2, made by a native and a non-native speaker of L2 (Pedersen 2009).

3   Collocations in the domain-specific language

The texts proposed as students' translation assignments (pieces of commercial correspondence, contracts and other documentation), in the same way as any language for special purposes, demonstrate a range of specific characteristics, such as a wide use of specialized terms and word combinations as well as stylistic features (Bowker & Pearson 2002: 25-27). Under this study though, we would like to concentrate on the second feature, i.e. on the particular way the words are combined within the domain of commercial Russian on the one hand, and on the (mis)collocations used by students in their translations, on the other. In this paper a statistic-based definition of collocations is adopted, according to which they refer to combinations of words with a significantly high co-occurrence frequency.

It is a well-known fact, that the use of collocations is a notorious problem of second language acquisition that persists even in advanced levels of language proficiency. Due to a transparent and comprehensible structure of many collocations, learners often don't identify them as "expressions" that are worth paying special attention to (Nesselhauf 2005: 252).

It was stressed by Sinclair that the meaning of the text is governed by the so-called idiom-principle (Sinclair 1991: 110). This characteristic of a text to be constructed out of a number of fixed chunks of words, is especially evident when it comes to official and bureaucratic writing. Commercial Russian in fact, is characterized by a high concentration of clichés and various types of fixed expressions. Let us consider some typical examples[1] of commercial correspondence:

(1) Компания «Мебельрос», динамично развивающееся предприятие, предлагает Вам долгосрочное сотрудничество, основанное на взаимовыгодных условиях.

(2) Руководство нашего предприятия проявило большой интерес к продукции Вашей компании, ознакомившись с её образцами на выставке «Сибирь 2015».

(3) Платёж производится в форме банковского перевода на основании выставляемых инвойсов, оформленных согласно международным торговым правилам.

These few examples demonstrate a variety of collocations of different structure: adjective + noun (*долгосрочное сотрудничество, взаимовыгодные условия*), adverb + verb (*динамично развиваться*), verb + (preposition) + noun (*проявлять интерес, ознакомиться с образцами, выставлять инвойс, оформлять инвойс*).

A special class of verb + noun combinations is constituted by so-called "stretched verb constructions" (Allerton 2002), in which the noun derivationally corresponds to a verb that expresses the same meaning as the whole noun + verb combination. The example of such a combination is presented in (3): *производить платёж – платить*. Use of "stretched verb constructions" is one of the most characteristic features of the official style of Russian language (e.g. *оказать поддержку, допускать ошибки, поднять вопрос*).

Thus, translations of texts of this genre and style pose additional difficulties for translators who should perfectly master the specific vocabulary and word combinations, or possess skills and strategies for efficacious research of the accurate equivalents of the target language. In the case of translation into a non-native language, the latter option, i.e. research skills, become most crucial in the professional figure of the translator and should therefore constitute a pivotal part of the translation teaching process.

4   Tools

The aim of any Translation class first of all, is to provide students with tools and methods enabling them to become more independent and confident translators (Woolard 2000: 33-36). So, which tools can be considered the most effective in order to enable our Italian-speaking students to successfully search for the right collocations of the domain-specific Russian language?

It is nothing new to state that traditional resources, such as paper mono- and bilingual dictionaries, due to their form, can provide only limited information about how the words are used in a language. We will return to this affirmation later in this paper.

Electronic on-line dictionaries on the other hand, have already become indispensable tools

---

[1] The examples are available in: Cadorin, Elisa, Kukushkina, Irina (2015) Il russo per l'azienda. Corrispondenza commerciale, documenti legali, dizionario commerciale. Milano: HOEPLI.

for translators, offering more up-to-date and quick-in-use lexical support. As for the Russian language, the most widespread electronic resource that is especially valuable for specialized translation is the multi-language Multitran electronic dictionary[2]. The main advantage of this resource is the fact that for each searched item, various field-specific equivalents are proposed. However, despite the large volumes of lexical items that this tool boasts, all the examples are restricted to short chunks of words, whereas a longer phrase-level context is not available.

In order to overcome the limitations of dictionaries, it is important to introduce linguistic corpora into the translation teaching process. Only they can provide such priceless information, as complete collocational and colligational word characteristics, frequency related and other quantitative data, as well as a wide range of examples of word use presented in its natural full context that is accessible for consultation (Aston 2000).

The most important reference corpus of the Russian language is undoubtedly the Russian National Corpus[3]. In fact, most university students of the advanced level are well aware of this tool and use it at various academic courses. At the moment, the main corpus counts 283 million words. However, the fact that fiction and historic texts constitute a considerable part of the corpus, whereas the texts of commercial and business spheres are scarcely represented, makes this tool hardly suitable for the needs of our course.

Huge volumes of up-to-date non-fiction textual data of Russian is contained in the Web-corpora, the largest of which is the Russian ruTenTen corpus that counts 14.5 billion words. The corpus is available within Sketch Engine[4] corpus manager that is by far one of the most powerful linguistic tools that allows exploring the words' combinability due to its unique Word Sketch functionality. The access to this corpus manager is not totally free though, as its academic license is limited to the EU area. We also believe that this tool is probably too large and sophisticated for classroom use, which was one of the reasons why Aranea Web Corpora Project[5] has been launched (Benko 2014).

The Aranea Project contains free gigaword Araneum Russicum Minus/Maius/Maximum corpora with the volumes of 120 million/1.2 billion/13.7 billion tokens respectively that are available within the NoSketchEngine environment. These corpora were designed for language and translation teaching purposes and that is why these tools were adopted in our classroom as the most suitable for the purposes of specialized translators training.

5   Corpora-based collocations mining

A selection of mis-collocations from our learner corpus is presented in Table 1 and confirms that the students' "collocational competence" (Hill 2000) definitely needs to be improved and in the first place, should be taught in a thorough way. The first column shows some passages of original Italian texts that were assigned for translation. In the second column a possible translation into Russian is proposed, whereas the third column demonstrates some students' translation attempts in their original (non-edited) form.

---

[2] Multitran is available on: *https://www.multitran.ru/*
[3] Russian National Corpus is available on: *http://www.ruscorpora.ru/*
[4] Sketch Engine is available on: *https://www.sketchengine.eu/*
[5] Aranea Web Corpora are available on: *http://aranea.juls.savba.sk/*

	Italian collocation	Russian collocation	Students' attempt
1	Vi preghiamo di **effettuare la consegna** ai nostri magazzini di Conegliano a mezzo camion.	Просим Вас **осуществить доставку** на наши склады в г. Конельяно грузовым автотранспортом.	Пожалуйста, **сделайте доставку** на наши склады в Конельяно наземным транспортом.
			Просим **доставить** нашим складам в Конельяно автотранспортом.
			Просим Вам **обеспечить доставку** в наших складах в Конельяно грузовиком.
2	I nostri cosmetici **offrono la massima affidabilità**.	Наша косметика **обеспечивает** максимальную **надёжность.**	Наша косметика **предлагает** самую высокую **надёжность**.
			Наши косметические изделия **предоставляют** высокую **надёжность**.
3	Vi abbiamo **riservato un vantaggio** eccezionale: uno sconto del 30% su tutti gli articoli acquistati.	Мы **предоставили** вам исключительное **преимущество**: скидку 30% на всю заказываемую продукцию.	Мы **сохранили** для вас исключительное **предложение**: 30% скидкой для всех покупаны изделия.
			Мы **предложили** Вам исключительное **преимущество**: 30% скидка на все приобретенные предметы.
			Мы **уготовили** для Вас исключительное **преимущество** - скидку в 30% на каждом покупанном продукте.
4	Attendiamo l'indicazione delle Vostre coordinate bancarie per poter **effettuare il pagamento** tramite bonifico.	Ждём указания ваших банковских реквизитов, чтобы мы смогли **произвести платёж** путём денежного перевода.	Мы ожидаем указания Ваших банковских реквизитов, чтобы **оплачивать** банковским переводом.
			Ожидаем ваши банковские реквизиты для того, чтобы **оплатить** через банковский перевод.
5	L'esportazione della merce sul territorio della dogana russa dovrà essere effettuata entro 90 giorni dalla data dell'**arrivo del pagamento** anticipato sul conto del Fornitore.	Вывоз товара на таможенную территорию РФ должен быть произведён в течение 90 дней с даты **поступления предоплаты** на счет Поставщика.	Экспорт Товара в территории русской таможнии должен быть выполниться не позднее 90 дней после даты **получения авансированной оплаты** на счёте Поставщика.
			Экспорт Товара на территорию российской таможни, должен быть произведен в течение 90 дней с даты **получения аванса** на счёт Поставщика.

Table 1. Examples from the learner corpus of translations.

One can see that some of the students' mis-collocations are produced due to the fact that they tend to translate single words from which the source Italian collocations are composed. For example, in the first example of Table 1 the source combination *"effettuare la consegna"* is translated as *"сделать доставку"*, in which *effettuare->сделать, consegna->доставка*.

So, the first activity proposed to our students is the following: after a brief review of the notion of collocation in class, students are asked to examine some typical commercial texts in Italian in which they should identify all the examples of useful (in their opinion) collocations and stretched verb constructions with the final discussion of the selected combinations.

This activity aims at raising students' awareness of collocations and is extremely useful as it makes students realize the importance of segmenting the source text not into single words, but into chunks of words (Lewis 2000; Hill 2000), that should be searched for and translated into the target language.

The awareness-raising activity concludes in a practical corpus-based task of searching for the right collocations in Russian, for which the Araneum Russicum corpora, as well as Concordance/Collocations/Frequency functionality of the corpus manager are introduced in class. Working with corpora in the classroom should demonstrate to the students, the techniques which will enable them to avoid some of the translation failures that are traceable in the learner corpus. Let us consider them in detail.

The failure (or unwillingness) to find an equivalent for the source stretched-verb constructions can be seen in examples 1 and 4 of Table 1, where some of the students' translations contain the corresponding single verbs, such as *"доставить"*, *"оплатить"*, that don't fit the syntactic structure of the phrase. The use of the Collocations function proves to be an easy way to identify the most frequent verbs to be used in a stretched-verb construction. For example, if we search for the *"Verb+платеж"* combinations in the Araneum Russicum Maius corpus, we will see that the most frequent verbs are *"произвести/осуществить/совершить"*:

	Cooccurrence count	Candidate count	T-score	MI	logDice
P | N произвести	102	90,088	10.088	9.873	5.190
P | N осуществить	91	51,666	9.532	10.511	5.810
P | N совершить	86	82,204	9.262	9.759	5.074
P | N обещать	83	61,778	9.102	10.120	5.426
P | N подтвердить	64	59,217	7.991	9.806	5.111
P | N провести	56	377,892	7.422	6.940	2.274
P | N внести	55	79,378	7.403	9.165	4.478

Picture 1. Left-side collocation candidates for the word *"платеж"* in Araneum Russicum Maius.

Secondly, the students' translation manifests the failure to identify an accurate collocation to fit the given context, which can be seen in example 5 of Table 1. The combination *"получение аванса/предоплаты"* has nothing wrong in itself, but it doesn't fit the right-hand context *"на счет поставщика"*. Picture 2 shows that limiting the right-hand context only to the examples containing the words *"на счет"* allows us to identify the most suitable collocation, i.e. *"поступление предоплаты"*.

	Cooccurrence count	Candidate count	T-score	MI	logDice
P | N поступление	11	47,800	3.316	13.217	2.913
P | N зачисление	4	7,663	1.999	14.398	4.090
P | N внесение	4	39,838	1.999	12.020	1.717

Picture 2. Left-side collocation candidates for the word "*предоплата*" in Araneum Russicum Maius.

Finally, the students not only have to find the right collocation, but also they have to be able to insert it correctly into the grammatical canvas of the phrase, which unfortunately they often fail to do. In example 3 of Table 1 students hesitate between the preposition *на*, that they are not sure which case should be followed by, and the preposition *для*, that corresponds to the source Italian preposition *per*. Consultation of the recent Italian-Russian Economic Dictionary[6] doesn't answer this question either.

If we examine the most frequent verbs that combine on the left side with the noun "*скидка*" (see Picture 3) we will reveal that the best collocate, that suits our context, is the verb "*предоставить*". After which the simplest way to define the preposition and the case to use with right-hand arguments (in our case, *на + accusative case*) is to explore the full contexts that lack in paper dictionaries, but provide precious information on the words behaviour (see Picture 4).

	lemma	Frequency	Items: 895 || Total frequency: 17,715
P | N	получить	3,250	
P | N	предоставляться	1,695	
P | N	получать	1,578	
P | N	действовать	1,450	
P | N	предоставлять	1,223	
P | N	делать	813	
P | N	сделать	628	

Picture 3. Frequency distribution of right-hand verb collocates in Araneum Russicum Maius.

Query **скидка** 100,201 > Positive filter (excluding KWIC) **V.*** 17,715 > Positive filter **предоставлять** 1,223 > Sort **Right** 1,223 (1.02 per million) 🛈

First | Previous Page 44 of 62 Go Next | Last   Concordance is sorted. Jump to: !⌄

luckyfamil...	торговым представителем ¶ Мы *предоставляем скидку* на закупки оптовыми партиями и обеспечиваем
kuhnibelar...	введены сертификаты, которые *предоставляют скидки* на замену фасадов. Кроме фасада, по желанию
zpford.ru ...	существенно дешевеет, так же мы *предоставляем скидки* на запчасти. ¶ Варианты оплаты: ¶ Оплата
516.ru	в течение февраля-марта *предоставляет скидку* на зимнюю одежду для животных в размере
newidey.ru...	фитнес-клубами, которые будут *предоставлять скидки* на игры в пейнтбольном клубе, при покупке
krb.in.ua ...	Стоматологическая клиника «Вилида» *предоставляет скидку* на израильские имплантаты зубов в период
tyumen.zar...	строительных компаний. - Мы *предоставляем скидку* на ипотеку от ведущих банков России. -
yarmetro.r...	клиентов банки-партнеры готовы *предоставлять скидки* на ипотечные займы, и наши брокеры сегодня

Picture 4. Extract from the concordance output for "*предоставлять скидку*" in Araneum Russicum Maius.

The goal of the above described, basic corpus-based activities is to demonstrate to students that they do have a tool that enables them to discover collocations and grammar patterns of the target language in a precise and totally independent way. From the pedagogic point of

---

[6] Hramova I.G., Lebedeva G.G. (2011) Novyj italjansko-russkij ekonomičeskij slovar'. Moskva: Filomatis.

view, it is this focus on skills developing and problem-solving procedures that makes the difference between translator training and translator education (Bernardini 2000) and should better prepare students for the passage into the professional world.

6   Data-driven word definition

Another corpora-based practice proposed to translation students is the "data-driven" words definition. This activity is based on a relatively old Firthian idea that "you shall know a word by a company it keeps" (Firth 1957: 179).
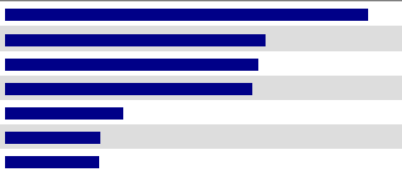
In case of words with similar meaning, it can be rather difficult even for a native speaker, to explain the difference in their behaviour: e.g. *продукт-продукция-изделие-товар-артикул, установка-устройство-оборудование-аппарат-приспособление-прибор, строительство-постройка-строение-сооружение-здание-корпус-застройка* etc. So, it is obvious that choosing between such alternatives is of great difficulty for a non-native speaker. Example 3 of Table 1 confirms this affirmation as the students hesitate exactly between the words "*изделия-предметы-продукты*", whereas the context determines the use of "*продукция-товары*".

During our translation course, the students were asked to use the Araneum Russicum corpora and their Collocations and Frequency functionality in order to define the differences in the use of the words *продукция-изделие-товар-продукт-артикул*, that are very frequent in the commercial domain. The goal of this activity is to draw the students' attention to the collocational fields of these words (Lewis 2000:13) and subsequently deduct the way they are distributed over various contexts.

For example, the adjective collocates of the word "*изделия*" suggest that the word is mostly used to designate some manufactured objects and that the preceding adjective usually indicates the type of source material (*ювелирные, железобетонные, металлические, меховые, кованые, керамические* etc.). On the other hand, the word "*продукция*" is mostly accompanied by adjective collocates designating the respective fields of production (*сельскохозяйственная, пищевая, косметическая, текстильная, печатная* etc.).

We would like to emphasise again that the results obtained with a web-corpus often do not have any reflection in dictionaries. The most frequent adjective collocate of the word "*продукт*" in Araneum Russicum Maius is "*программный*" (see Picture 5), a combination that can be translated into English as *software*. Neither the most recent technical and economic bilingual dictionaries [7], nor the electronic Multitran dictionary register this collocation.

---

[7] The following dictionaries were consulted: Cadorin E., Kukuškina I. (2017) Dizionario tecnico russo. Milano: HOEPLI; Hramova I.G., Lebedeva G.G. (2011) Novyj italjansko-russkij ekonomičeskij slovar'. Moskva: Filomatis.

lemma	Frequency	Items: 8,011 || Total frequency: 139,812
P | N программный	10,806	
P | N пищевой	7,753	
P | N новый	7,549	
P | N молочный	7,384	
P | N натуральный	3,501	
P | N конечный	2,818	
P | N качественный	2,787	

Picture 5. Frequency distribution of right-hand adjective collocates "*продукт*" in Araneum Russicum Maius.

We should stress that the proposed activities are of high importance from the "psychological" point of view as well. The examples above demonstrate situations in which the teacher is unlikely to come up with a good explanation for every single word choice. So we suppose that rather than trying to provide students with explanations, it is much more effective if teachers "come down from their pedestal of a native-speaker" (Kiraly 2000) and search for the answers in the corpora data along with the students. It is this co-working on equal terms with learners that has an important benefit on self-esteem and confidence which is so crucial when translating into a non-native language.

7   Feedback and conclusion

In this paper, we tried to demonstrate an approach that sets as the final goal of translation teaching the acquisition of skills enabling students to work in an autonomous and more confident way. In our opinion, the concordancing activities based on the huge web-corpora should become regular practice in both the classroom and future professional activity. The result of the regular use of Aranea web-corpora in the training process of Specialized Translation is tangible in the students' growing competence of domain-specific Russian, more accurate translations of specialized texts into a non-native language and "a shift toward a more student-centred classroom" (Kiraly 2000).

In conclusion, we would like to share the students' enthusiastic feedback collected at the end of the course. The Aranea web-corpora were unanimously recognised as a useful tool for a translator. The students have demonstrated that the use of this tool becomes especially efficacious when it comes to distinguishing between synonymous words, collocation and terminology search. The students also admitted that the use of this tool makes them more confident when they have to work into a non-native language and that they are going to adopt it in their future professional practice.

8   References

Allerton, D. J. (2002) Stretched verb constructions in English. London/New York: Routledge.

Aston, G. (2000) I corpora come risorse per la traduzione e per l'apprendimento. In S. Bernardini, F. Zanettin (eds.) (2000) I corpora nella didattica della traduzione. Atti del Seminario di Studi Internazionale. Bologna: CLUEB, pp. 21-29

Benko, V. (2014) Aranea: Yet Another Family of (Comparable) Web Corpora. In P. Sojka, A. Horak, I. Kopeček, K. Pala (eds.): Text, Speech and Dialogue. 17th International Conference, TSD 2014, Brno, Czech Republic, September 8–12, 2014. Proceedings. LNCS

8655. Switzerland: Springer International Publishing, pp. 257–264.

Bernardini, Silvia (2000) I corpora nella didattica della traduzione: dall'addestramento alla formazione, In: Silvia Bernardini, Federico Zanettin (eds.) (2000) I corpora nella didattica della traduzione. Atti del Seminario di Studi Internazionale, Bologna: CLUEB, pp.81-100.

Bowker, L., Pearson, J. (2002) Working with Specialized Language: a practical guide to using corpora. London/New York: Routledge.

Firth, J. R. (1957) A synopsis of linguistic theory, 1930-55. In F.R. Palmer (ed.) Selected Papers of J.R. Firth 1952-59. London: Longmans, pp.168-205.

IAPTI (2015) Translation into a non-native language. IAPTI's Ethics Committee. Accessed at: https://www.iapti.org/files/surveys/2/IAPTI_non-native_report.pdf [26/06/2018].

Hill, J. (2000) Revising priorities: from grammatical failure to collocational success. In: M. Lewis (ed.) (2000) Teaching collocation: further developments in the lexical approach. England: Language Teaching Publications, pp. 47-69.

Kiraly, D. C. (2000) Translation into a non-mother tongue: From collaboration to competence. In M. Grosman, M.Kadric, I.Kovacic, M.Snell-Hornby (eds.) (2000) Translation into Non-Mother Tongues in Professional Practice and Training. Tübingen: Stauffenburg Verlag, pp.117-123.

Lewis, M. (2000) There is nothing as practical as a good theory. In M. Lewis (ed.) (2000) Teaching collocation: further developments in the lexical approach. England: Language Teaching Publications, pp.10-27.

Nesselhauf, N. (2005) Collocations in a Learner Corpus. Amsterdam/Philadelphia: John Benjamins.

Pederson, V. (2009) Translation into L2 – In practice, and in the classroom. In M. Grosman, M. Kadric, I. Kovacic, M. Snell-Hornby (eds.) (2009) Translation into Non-Mother Tongues in Professional Practice and Training. Tübingen: Stauffenburg Verlag, pp. 109-114

Sinclair, J. (1991) Corpus, Concordance and Collocation. Oxford: Oxford University Press.

Woolard, G. (2000) Collocation – encouraging learner independence. In M. Lewis (ed.) (2000) Teaching collocation: further developments in the lexical approach, England: Language Teaching Publications, pp. 28-46