

## Creating a bilingual dictionary of collocations: A learner-oriented approach

Sonia Vaupot\*

Department of Translation Studies, University of Ljubljana, (Aškerčeva 2, 1000 Ljubljana), Slovenia

### ABSTRACT

Considering the lack of specialised dictionaries in certain fields, a creative way of teaching through corpora-based work was proposed in a seminar for Master's students of translation studies held at the University of Ljubljana, Slovenia. Since phraseology and terminology play an important role both in specialised translation and in the learning path of students of translation studies, this article presents an active approach aimed at creating an online lexicographic resource in languages for specific purposes by using the didactic tool and database ARTES (Aide à la Rédaction de TExtes Scientifiques/Dictionary-assisted writing tool for scientific communication) previously developed at the Université de Paris in France. About thirty Slovene students enrolled in the first year of their Master's programme have been participating in the bilateral project since 2018. The aims of such an activity are multiple: students learn in a practical way how to compile corpora from the Internet, using the online corpus software Sketch Engine, to find similar linguistic constructions in the source and target languages. They also learn to create an online bilingual phraseological and terminological dictionary to facilitate the translation of specialised texts. In this way, they acquire skills and develop some knowledge in terms of translation, terminology, and discourse phraseology. The article first describes the ARTES online database. Then, we present the teaching methodology and the students' work, which consists of compiling corpora, extracting and translating collocations for the language pair French-Slovene, and entering them in the ARTES database. Finally, we propose an analysis of the most frequent collocation structures in both languages. The language pair considered here is French and Slovene, but the method can be applied to any other language pair.

**Keywords:** Bilingual dictionary; collocations; corpora; language learning; specialised translation

**First Received:**

14 July 2020

**Revised:**

7 November 2020

**Accepted:**

28 December 2020

**Final Proof Received:**

21 January 2021

**Published:**

31 January 2021

### How to cite (in APA style):

Vaupot, S. (2021). Creating a bilingual dictionary of collocations: A learner-oriented approach. *Indonesian Journal of Applied Linguistics*, 10(3), 762-770.  
<https://doi.org/10.17509/ijal.v10i3.31888>

### INTRODUCTION

Idioms and collocations belong to the set phrases of a language. Collocations, arbitrary and recurrent word combinations, are expressions whose importance in language has been increasingly noted in recent years. They are also referred to as prefabricated units, phraseological units, (lexical) chunks, prefabs, multi-word units, etc. (Wray, 2002). The collocations can be divided into two groups: grammatical collocations and lexical collocations. Grammatical collocations consist of a dominant word (noun, adjective, or verb)

and a dependent word (preposition or a grammatical structure such as an infinitive or clause). Some examples of grammatical collocations are, for instance, *account for*, *by accident*, *to be afraid that*. Lexical collocations contain various combinations of two equal words (some combinations contain nouns, verbs, adjectives, and adverbs): for example, *inflict damage*, *extreme poverty*, *directly concerned*.

The collocations can be a source of difficulty for non-native speakers of a language (Leed & Nakhimovsky, 1979; Mc Alpine & Myles, 2003). A

common phrase typically used in the target language often has to be learned verbatim or cannot be translated on a word-by-word basis. Various studies have been conducted to determine whether non-native speakers have problems decoding or encoding collocations, or to determine the extent to which dictionaries help learners use collocations. Moreover, lexical errors in general and collocational errors in a foreign language are common due to a linguistic calque. The collocations do not necessarily have a literal equivalent in another language (Siepmann, 2006) and learners are often not aware of, or familiar with, the collocate. The expressions that cause the most problems result from the association of words that do not belong together in native language usage, whose translation is context dependent, or combinations of basic items that all learners should be familiar with (Binon & Verlinde, 2003). The collocations obviously play an important role in language learning, and are essential for fluency in spoken and written language. They involve comprehension, so that the learner understands the meaning of a passage of text without having to pay attention to every word (Hunston & Francis, 2000, p. 270) and they fulfil “[...] the desire to sound [and write] like others” (Wray, 2002, p. 75). Knowing how to use collocations is thus essential for language learners, and there is agreement that collocations need to be taught (Nesselhauf, 2003).

In recent years, several approaches to language teaching have been developed that place collocations at the centre of teaching: Lewis’ Lexical Approach (1997, 2000), Nattinger and De Carrico’s Lexical Phrases Approach (1992); the Distributional Approach to define collocations (Granger & Paquot, 2008), while numerous scholars have published papers on collocations (Cavalla, 2018; Tutin & Grossmann, 2002). Lewis (2000) proposed a new lexical approach that focuses on teaching lexical chunks. He argues that language consists of combined chunks that make up a coherent text, and that we should raise learners’ awareness of collocation. He suggests “[...] we now recognize that much of our vocabulary consists of prefabricated chunks of different kinds. The single most important kind of chunk is collocation. Self-evidently then, teaching collocation should be a top priority in every language course.” (ibid., p. 8). This is also the opinion of Nattinger and De Carrico (1992, p. 32), who claim that:

It is our ability to use lexical phrases that helps us to speak with fluency. This prefabricated speech has both the advantages of more efficient retrieval and of permitting speakers (and learners) to direct their attention to the largest structure of the discourse, rather than keeping it narrowly focused on individual words as they are produced.

Other linguists (Firth, 1957; Halliday & Hasan, 1976; Pecman, 2012; Peeters, 2019) attach great

importance to the fact that collocation can contribute to textual cohesion:

The cohesive effect of such pairs [laugh ... joke, blade ... sharp, ill... doctor] depends not so much on any systematic relationship as on their tendency to share the same lexical environment, to occur in COLLOCATION with one another. In general, any two lexical items having similar patterns of collocation - that is, tending to appear in similar contexts - will generate a cohesive force if they occur in adjacent sentences. (Halliday & Hasan, 1976, p. 285-286)

More recently, the issue of the accessibility of phraseological information in dictionaries has also been raised (Herbst & Mittmann, 2008), and research has focused on aspects of coverage, such as the number of phraseological units listed in dictionaries (Götz-Votteler & Herbst, 2009).

From a more didactic point of view, the Common European Framework of Reference for Languages (CEFR, 2001; cf. Council of Europe), for its part, briefly defines phraseological units. Chapter 5, entitled “The user/learner’s competences”, divides linguistic competences into six types, but only two of them concern collocations. First, the lexical competence, which consists, among other things, of fixed expressions (sentential formulae, proverbs, relict archaisms), phrasal idioms (semantically opaque, frozen metaphors, intensifiers), fixed frames or meaningful sentences, but also other fixed phrases (phrasal verbs, compound preposition), and fixed collocations consisting of words that are regularly used together (e.g., to make a speech or to make a mistake). Secondly, semantic competencies - including lexical semantics - deal with issues of word meaning, i.e., the relation of the word to the general context, and with inter-lexical relations, which include collocations. In order to best develop learners’ linguistic competence in relation to vocabulary, the CEFR (ibid., p. 150) recommends developing vocabulary by explaining and training the use of lexical structure. The New Descriptor of the Companion Volume (2018, p. 133) adds that vocabulary control concerns the learner’s ability to choose an appropriate expression: “As competence increases, such ability is driven increasingly by association in the form of collocations and lexical chunks, with one expression triggering another” (ibid., p. 181) in the written assessment grid. Teaching phraseology is not generally recognized by the CEFR, which does, however, require lexical competence to be mastered by the end of training.

On the other hand, since the 1980s, corpus linguistics has opened up new possibilities for the study of language in general. Some methods have been proposed for the automatic extraction of collocations from text corpora. Collocation encoding can indeed provide useful lexical information about the conventionalities of languages, and such resources can be useful for language learners or non-

native speakers. Moreover, encoding collocations in a terminological database that contains the terminology of a particular scientific or specialised field as well as the most common collocational patterns in which that terminology occurs can provide useful lexical information about the conventionalities of languages for specific purposes (Pecman, 2007, 2012).

Considering the lack of specialised dictionaries in certain fields, especially for the language pair French-Slovene, a creative way of teaching through corpora-based work was proposed in a seminar for master's students of translation studies (University of Ljubljana, Slovenia). Since phraseology and terminology play an important role both in specialised translation and in the learning path of students of translation studies, this article presents an active approach aimed at creating an online lexicographic resource in languages for specific purposes. The method is based on a project carried out by researchers from the Faculty of Arts, University of Ljubljana, Department of Translation Studies, on the Slovene side, and the research team from the Center for Linguistics, Interlanguage, Lexicology, English, and Corpus Linguistics (CLILLAC) at the Université de Paris (formerly Paris Diderot University, also known as Paris 7) on the French side. The project involves the development of Slovene-French (and French-Slovene) terminology and phraseology resources for specialised translation. It requires the transfer of skills related to the processing of specialised lexicography and lexicology, with the aim of providing the necessary basis for collaboration in common language resources within the online ARTES database (*Aide à la Rédaction de Textes Scientifiques/Dictionary-assisted writing tool for scientific communication*). The ARTES dictionary is simultaneously a teaching tool for training future translators in terminology and phraseology, and a linguistic resource, bringing together a lot of useful information for specialised translation purposes. The aims of such an activity are multiple: students learn to compile comparable corpora from the internet to find similar linguistic constructions in the source and target languages. They also learn to create an online bilingual phraseological and terminological dictionary to facilitate the translation of specialised texts. In this way, they acquire skills and develop some knowledge in translation, terminology, and discourse phraseology. The article first describes the ARTES online database. Then, we present the teaching methodology and the students' work, which consists of extracting and translating collocations for the language pair French and Slovene. Finally, we propose a synthesis and an analysis of the most frequent collocation structures in both languages. The language pair treated here is French and Slovene, but the methodology can be applied to any other language pair.

## **METHOD**

The ARTES database is designed for the creation of multilingual and multi-domain resources. It is a tool that helps users write or translate texts for Specific Purposes. It was developed in 2010 by the French research team from the CLILLAC-ARP research centre and the EILA department of Paris-Diderot University. ARTES is also used as a didactic tool for teaching terminology and phraseology to translation students. With the database ARTES, it is possible to look up terms from different subject areas and find out their most frequent contexts of use, their terminology or phraseology, as well as the search for common expressions used in different specialised discourses. The tool has a dictionary of terms with definitions, useful contexts, collocations, synonyms, and finally the translations of terms. Via the dictionary of expressions, it is possible to learn more about transdisciplinary phraseology, and find out the role and translations of different transdisciplinary lexico-grammatical structures. It is also possible to use multi-criteria search functions. ARTES is designed to allow multilingual external collaboration. It has been adapted to about 50 languages (Kübler & Pecman, 2012). The database was set up precisely with the aim of enabling external collaboration and has been adapted to Slovene (see Figure 1).

Access to the ARTES online lexicographic database was provided by the French research team. Sources were, and still are, collected by students each year and entered directly into the ARTES dictionary. This collaboration helped to create a Slovene and French corpus, and to provide users with the necessary skills to compile a specialised online dictionary Slovene-French within the ARTES database.

## **Participants**

To create monolingual or bilingual dictionaries and terminology databases, translators can extract a large amount of data from the corpora. Comparable corpora consisting of authentic texts have become tools in the creation of bilingual dictionaries. The value of using corpora, especially for specialised translation, is well-established (Kübler, 2011; Morin & Daille 2006, 2012). Since corpora for language pairs that do not include English are rarer, the first step is to assemble a specialised corpus. In a monolingual context, collocations are recognized based on recurrence in many texts, which can only be done with the help of large text corpora.

About thirty Slovene students enrolled in the first year (MA1) of the master's program in Translation Studies have been participating in the bilateral project since 2018. The students use the pre-designed database/dictionary ARTES to encode phraseological information through a corpus-based study in the field of diplomacy and international relations (2018-2019). Thus, ARTES is fed by students' work each year to develop a Slovene-

French language combination with the aim of creating a phraseological database in specialised fields using corpus-based resources. Therefore, students are asked to create two comparable corpora: a French corpus and a Slovene corpus. They first define the domain and then build up two comparable corpora in French and in Slovene in the microdomain of diplomacy or international relations (Udovič,

2016). The specialised fields that have been covered by their corpora are, for example, humanitarian diplomacy, cultural diplomacy, economic diplomacy, Brexit, human rights, political speeches, foreign policy, imperialism. The corpora obtained in this way are variable, and contain between 250,000 and 600,000 words, depending on the subject area.

Figure 1

The ARTES Dictionary Interface (<https://artes.app.univ-paris-diderot.fr/artes-symfony/web/app.php>)



After building the corpora (Slovene and French), students need to carry out a phraseological project using the ARTES database, including collocations for the two source languages and their equivalents in the target languages. Since there are already some studies on the problem of translating collocations from a specialised corpus (Kübler, 2003; Pecman, 2007), which have led to a separation into specific collocations (associated with terminology) and generic collocations (associated with discourse), our attention has focused on the extraction and entry of generic collocations. Generic (i.e., domain-free) collocations are associated with discourse functions, and their usage cannot be ascribed to a specific domain (for example, these findings may be the first to be described), and to the dominant discourse type: for instance, scientific, technical, administrative, socio-economic, political. (Kübler & Pecman, 2012, p. 202). Thus, the database provides users with a valuable resource for reading, writing, or translating specialised texts or genres.

As in the field of terminology, text-based approaches or lexicography (Kübler & Pecman, 2012; L'Homme, 2019), the search for generic collocations can be based on the content of specialised texts. The process of creating these resources consists of several phases: (a) students build specialised corpora; (b) they select from their corpus the most frequent and interesting generic collocations for translation purposes; (c) they

manually add the generic collocations and their context to the ARTES database, (d) they identify the equivalents of the generic collocations, (e) they upload their translations into ARTES, considering the context. In addition to building a specialised corpus and entering the generic collocations into ARTES, students present the results of their research in a seminar paper, which they submit at the end of the semester. They receive detailed instructions on how to do this at the beginning of the semester.

### Collocation extraction

It is well known that the web is a mine of language data that is easily accessible. It is also a viable source of corpora created ad hoc for a specific purpose. In our case, we use the Sketch Engine (<https://www.sketchengine.eu/>), to create corpora and compile phraseological databases. Sketch Engine is an online corpus software with a variety of features that can be used for pedagogical purposes. Using this software allows us to automate the process of searching for reference texts on the Internet and compile them into a single corpus. One can quickly build a relatively large corpus. Therefore, it is a useful tool for translators and students, and has been used in translation or terminology classes to build corpora of different sizes and specialization. Thus, the spectrum of phraseological data in the context of languages for special purposes provides students or

other users with a valuable resource for reading, writing, or translating specialised texts or genre.

The methodology used by the students is based on automatic collocation extraction using the Sketch Engine tool. One of the functions is Word Sketch for extracting collocations in a range of grammatical patterns. The results are organized into grammatical relationships, such as words that serve as the object of the verb, words that serve as the subject of the verb or words that modify the verb. For extracting the generic collocates of the vocabulary, students identify the collocations that occur in the corpus. They automatically extract the collocates of the selected query lemmas in the corpora they have built up. As they are interested in collocations of generic vocabulary used in diplomacy and international relations from Slovene, they extract the corresponding collocations in French, and vice versa. Consequently, the methodology for extraction consists of lemma selection, collocation extraction and collocation filtering.

Before they started the collocation extraction process, the students selected the query lemmas for which the collocates were to be extracted. They selected the most frequently occurring collocate in one of the corpora. They then identified the lemmas that occur in both corpora. The selection was also based on the comparison of frequencies between the two corpora. To identify the generic collocations, they used the Collocation Function from the installation of the Sketch Engine (Kilgarriff et al., 2004). They extracted the lemma and the part of speech of the collocate in the Sketch Engine, as well as information about the frequency of the collocation. They extracted collocates in adjacent positions, i.e. immediately preceding or following the lemma. Next, the extraction depended on the input lemma type.

Consequently, the extraction of generic collocations is based on the comparison of collocations in both corpora, and the resulting collocation lists contain the collocations in each of the two corpora. For the selected query lemmas, students extracted five collocations in Slovene and five collocations in French, for a total of ten collocations. The process of creating these resources thus involved the manual entry of Slovene or French collocations, respectively, into the ARTES database, followed by the identification of their equivalents, which in turn were also added to the database. These resources provide a valuable insight into measuring the current state and trends in Slovenia, including in the field of translation.

## RESULTS AND DISCUSSION

A corpus-based lexical analysis allows to reveal, among other things, collocation and phraseological patterns. In this way, the meaning of a word is inferred based on its prototypical use in the concordances (Endarto, 2020).

Consideration of phraseology is one of the approaches to the data proposed in ARTES, which provides, in our case, an onomasiological approach to collocations common to a variety of languages for special purposes (LSP) discourses, and serves as a tool for scientific drafting (Pecman, 2007, 2008). Some labelling tables contain open-class type values, such as the discourse functions tables that offer about eighty classes for categorizing generic collocations according to their meaning or function in LSP discourses. They can be modified or completed according to the results of the research conducted. Domain-free or generic collocations are associated with discourse functions. A brief selection of generic collocations attributed to different discourse functions is given in Table 1.

**Table 1**  
*Generic Collocations and Discourse Functions*

<b>Generic collocation (Slovene or French)</b>	<b>Associated discourse function</b>
<i>odnosi so ključnega pomena za (relationships are crucial for)</i>	Describing and analysing observed data or phenomena
<i>verjamemo v prihodnost (we believe in the future)</i>	Expressing effect, cause, consequence
<i>nous nous intéressons à (we are interested in)</i>	Announcement of the subject of the current section

Some other labelling tables contain closed-class type values, such as the grammatical functions table that provides ninety categories that are linked to the table of terms.

### Semantic prosody

Semantic prosody has been a field of linguistics and lexicographical exploration for more over than two decades. For Louw (1993, p. 157), the term itself describes “the consistent aura of meaning with which

a form is imbued by its collocates”. According to Sinclair (2004, p. 23), it is an “attitudinal or pragmatic meaning” that exists alongside “the familiar classificatory meaning of the regular dictionary”, i.e. denotation. Some authors also reserve the term semantic prosody “for the attitudinal discourse function of a larger unit of meaning, with the word at its core” (Louw, 2000; Siepmann, 2005, 2006; Sinclair, 2004).

One approach that the ARTES dictionary highlights is the need to consider the fact that semantic prosody and semantic preferences are particularly useful for understanding the discourse and structure of the lexicon. For Sinclair (1996, p. 87), semantic prosody lies on the "...pragmatic side of the semantics/pragmatics continuum". From a semantic perspective, collocation is represented by semantic preference and semantic prosody, both of which describe the significant co-occurrence of a word with a group of other words. Semantic preference deals also with a semantic set of collocates that share part of a set of semantic features (Kübler & Pecman, 2012, p. 188).

First, it is noticeable that the prevalence of extracted lexical collocations compared to grammatical collocations. Moreover, most of the collocations extracted from the students' corpora are neutral or positive. The Slovene verbs that are most used with collocations are also neutral or positive: *biti* (*être*), *doseči* (*atteindre, parvenir à*), *imeti* (*avoir*), *izvajati* (*exécuter, mener faire, effectuer*), *sprejeti* (*prendre, passer, adopter*), *zagotoviti* (*apporter, créer, assurer, fournir*). Only a few

categories of expressions can be identified as negative: *izvajati pritisk* / *faire pression sur*; *napovedati vojno* / *déclarer la guerre*; *obrniti hrbet* / *tourner le dos*; *povzročiti padec* / *entraîner une baisse*; *pranje denarja* / *blanchiment d'argent*. The results of the study thus confirm the interdependence between lexicon and grammar. Indeed, knowledge of grammatical and syntactic regularities makes it possible to identify, in the lexical productions, what belongs to productive mechanisms. The grammar and lexicon, then, cannot be separated in lexicography. Rather, they merge into each other (Willis, 1990): the lexical meaning is actualized in specific syntactic patterns and in typical contexts of occurrence. The principle of lexico-grammar is also one of the foundations of foreign language didactics.

### Collocations selected according to the frequency of use

Among the grammatical, mainly syntactic functions, generic collocations belonging to the following categories were recorded and translated by the students:

**Table 2**

*Collocations and Their Constructions in the Source Language and the Target Language*

	Slovene	French	Examples
1	Vb + N	Vb + N	<i>adoptir les mesures/ sprejeti ukrepe</i> ( <i>adopt the measures</i> )
2a	Adj + N	N + Adj	<i>le traitement équitable/ pravična obravnave</i> ( <i>fair treatment</i> )
2b	Adj + N	N + prep (de) + N	<i>tiskovna konferenca / conférence de presse</i> ( <i>Press conference</i> )
3	N + N	- N + prep + N	- <i>varstvo potrošnikov / protection du consommateur</i> ( <i>consumer Protection</i> )
		- N + Adj	- <i>združitev družine / regroupement familial</i> ( <i>family reunification</i> )
		- Vb + prep + N	- <i>odpoved imuniteti / renoncer à l'immunité</i> ( <i>waive immunity</i> )
		- nominal construction	- <i>sklenitev sporazuma / la mise en place de l'accord</i> ( <i>conclusion/implementation of the agreement</i> )

It can be confirmed that it is possible to find similar linguistic constructions in the source language and in the target language by using comparable corpora. In (1), the grammatical categories are the same in both languages; in (2a) the Slovene structure (Adj.+N) corresponds mainly to the French inverted grammatical structure (N+Adj), but some collocations in this category (2b) are also translated by a verbal prepositional construction. As we can see in (3), this category can be translated in many ways.

### Comments on translation results

In general way, it can be seen that students translated generic collocations by using the same or equivalent grammatical structures, and close or semi-equivalent grammatical structures.

As can be seen in Figure 2, the equivalent grammatical structures in French and Slovene are as follows:

- vb + N (frequency of use: 86%): *imeti pogum za / avoir le courage de* (*have the courage to*)
- vb + prep + N (frequency of use: 10%): *imeti na zalogi / avoir en stock* (*have in stock*)
- vb + N + prep (frequency of use: 4%): *izvajati pritisk na / faire pression sur* (*put pressure on*)

The semi-equivalent grammatical structures are as follows in the Table 3.

Figure 2

Equivalent Grammatical Structures between French and Slovene

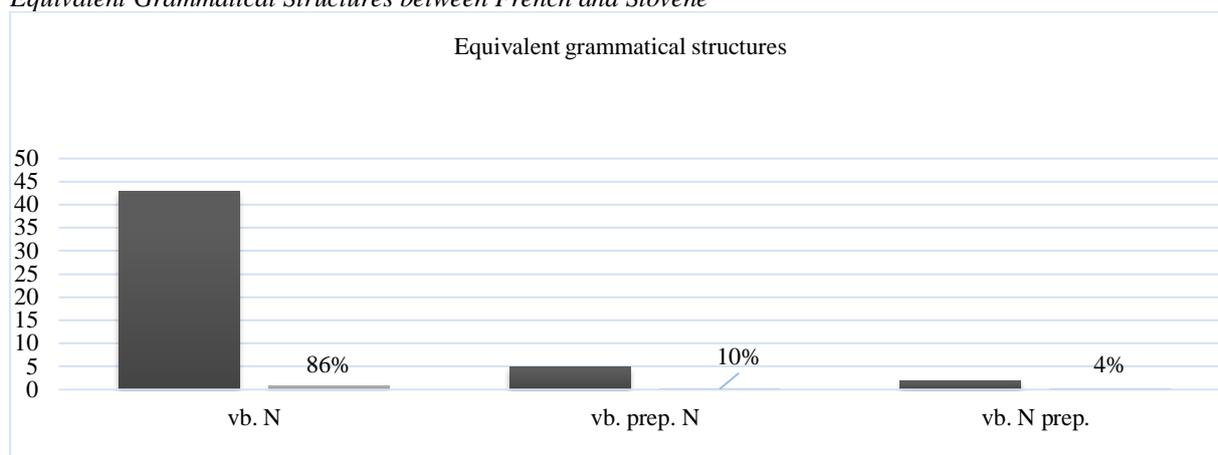


Table 3

Semi-Equivalent Grammatical Structures between French and Slovene

	Slovene	French		Examples
1	Adj. + N	N + Adj.	39 %	<i>diplomatsko predstavništvo / mission diplomatique (diplomatic mission)</i>
2	Adj + N	N + prep + N	17 %	<i>izredne razmere / situation d'urgence (emergency situation)</i>
3	vb + Adj.+ N	Vb + N + Adj	12 %	<i>imeti ključno vlogo / jouer un rôle majeur (play a key role)</i>
4	N + N	Vb + N	12 %	<i>izgrajevanje ugleda / promouvoir l'image (building reputation/promote the image)</i>
5	N + N	N + prep + N	12 %	<i>varstvo potrošnikov / protection du consommateur (consumer protection)</i>
6	prep + N	prep + N + prep	8%	<i>pod vodstvom / sous la direction de (under the leadership of)</i>

In Table 3, it can be noted that the main Slovene grammatical structure (Adj + N), entered into the ARTES database, is reversed in French (N + Adj). We can also note that the Slovene grammatical structure (N + N) includes the Slovene verbal noun (in Slovene, *glagolnik*), which is a nominal form consisting of an infinitive verb ending with *-anje*. This category is used to express a state or an action. Most students translated this form into French by using a verb.

Therefore, it is noticeable that the notion of equivalence represents a certain homogeneity between the original collocation and its translation. The number of identical verbal collocations is relatively large in contrast to nominal collocations, which are translated in various ways. The translation of collocations requires the translator to master the collocation systems of the languages involved in the translation. In the absence of such mastery, collocations can become real pitfalls in translation. However, the fact that comparable corpora were used seems to have made the task of translation easier for the students. Indeed, the students noted in their

presentation file that they had not encountered any major problems in translation. To translate generic collocations, they used their corpus and the following online dictionaries: Linguee, Glosbe, Iate, Pons, Reverso, Evroterm, Termium, WordReference, Fran, Larousse. Using these, they selected the best translation according to the context. However, they mentioned some linguistic problems, especially the problem of alignment; the difference in grammatical forms (nominal, verbal) and the position of adjectives, which are different in the two languages. On the other hand, they sometimes found it difficult to find an equivalent generic collocation, which is why they sometimes used the simple form of the verb. For example, the expression “*mettre en place*” is common in French and can occur in different contexts. However, in the absence of an exact equivalent in Slovene, the collocation was translated with the simple verb “*vzpostaviti*” and not with an equivalent collocation. The use of the passive form also raised some difficulties. According to Slovene grammar, the passive should be avoided: Thus, the collocation “*être compris comme*” is translated with

the active form in Slovene “*Razumljen kot*”. It was also mentioned that the use of bilingual dictionaries is not very helpful in the search for equivalence, so it was considered important to refer to a corpus. Finally, the translation of some collocations required more detailed research for some students, which depended on the subject area. Indeed, the field of diplomacy presents some translation problems due to diplomatic conventions, and the (non)translation of certain terms: for instance, the expression “*le bout de papier*” cannot be translated because it is specific to the field of diplomacy and is used to describe a relatively informal communication or record of a meeting.

Although the equivalence between the constructions in comparable corpora may not be complete, it can be confirmed that there is a sufficient similarity between the resources available in the two languages. The study gives an idea of possible solutions for non-literal translation. In this sense, the ARTES tool is more than just a dictionary: it can point translators to potentially good and contextually appropriate suggestions.

## CONCLUSION

In this article, a creative method for guiding students on subject phraseology through corpora-based work on a selected domain was presented. The research study focuses on the needs of Slovene students. It describes part of the process through which an online bilingual LSP dictionary was created. There are some advantages and challenges associated with this method. The students build a specialized corpus using the Sketch Engine corpus manager and text analysis software, which requires technical skills. Once the corpora are built, they must find and extract the specialized term or collocation in the source language, propose an equivalent term or collocation in the target language based on co-occurrences, and validate it against a comparable context. The students make a selection based on the frequency of occurrence in the corpus and the difficulty of translating a term or collocation. It may be that translation equivalence does not exist, but they must solve the translation problem. Then they have to integrate the term or collocation into the database. They learn that there are clear criteria that the lexicographer can follow in compiling a bilingual dictionary. Thus, both linguistic (lexical, semantic, grammatical, etc.) and lexicographic knowledge is required for the analysis, morphological extraction, translation and integration of terms and collocations into the ARTES database.

With such an active lexicographic approach, students learn how to concretely use corpora to create a Slovene-French online dictionary. The overall evaluation of this project is very positive, as students have made progress on several levels: in the creation of specialised corpora in the field of diplomacy and

international relations, in the extraction and input of generic collocations, and in translation into their native language, Slovene, and into a foreign language, in our case French. Moreover, the results are linked to the students' work and profile. They improved their language level and their knowledge of corpus linguistics. They made decisions independently or, when necessary, with the help of the teacher to overcome the student's problems. In conclusion, they found that the creative way of teaching through corpus-based work is an interesting and useful method, but it was not even easy to use. Since the teaching method covers several areas, it is most suitable for advanced students who are willing to invest in their work. In addition, another teaching framework or path using the ARTES database can also be based on terminology management, collaboration with experts and analysis of translations, which can also be used to provide research material on translation problems.

## REFERENCES

- Binon, J., & Verlinde, S. (2003). Les collocations : clef de voûte de l'enseignement et de l'apprentissage du vocabulaire d'une langue étrangère ou seconde. *La Lettre de l'AIRDF*, 33(2), 31-36.
- Cavalla, C. (2018). Exemple d'enseignement de la phraséologie transdisciplinaire à l'aide de corpus numériques en FLE. *La Lettre de l'AIRDF*, 64(1), 43-47.
- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge University Press.
- Council of Europe (2018). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume with New Descriptors*. Cambridge University Press.
- Endarto, I. T. (2020). A corpus-based lexical analysis of Indonesian English as a new variety. *Indonesian Journal of Applied Linguistics*, 10(1), 95-106.
- Firth, J.R. (1957). *Papers in linguistics, 1934-1951*. London, New York, Oxford University Press.
- Götz-Votteler, K. & Herbst, T. (2009). Innovation in advanced learner's dictionaries of English. *Lexicographica*, 25, 47-66.
- Granger, S., & Paquot, M. (2008). 2. Disentangling the phraseological web. *Phraseology*, 27-49.
- Halliday, M., & Hasan, R. (1976). *Cohesion in English (English Language Series; No. 9)*. London: Longman.
- Herbst, T. & Mittmann, B. (2008). Collocation in English dictionaries at the beginning of the twenty-first century. *Lexicographica. International Annual for Lexicography*, 103-19.

- Hunston, S., & Francis, G. (2000). *Pattern Grammar: A corpus-driven approach to the lexical grammar of English (Studies in Corpus Linguistics)*. John Benjamins Publishing Company.
- Kilgarriff, A., Rychlý, P., Smrz, P., & Tugwell, D. (2004). The Sketch Engine. *Proceedings of the Eleventh EURALEX International Congress, 1*, 105-116.
- Kübler, N. (2003). Corpora and LSP translation. *Corpora in translator education*, 25-42.
- Kübler, N. (2011). Working with different corpora in translation teaching. *New Trends in Corpora and Language Learning*, 62-80.
- Kübler, N., & Pecman, M. (2012). The ARTES bilingual LSP dictionary: From collocation to higher order phraseology. *Electronic Lexicography*, 187-210.
- Leed, R. L., & Nakhimovsky, A. D. (1979). Lexical Functions and Language Learning. *The Slavic and East European Journal*, 23(1), 104-112.
- Lewis, M. (1997). *Implementing the Lexical Approach*. Putting Theory into Practice. Hove, England: Language Teaching Publications.
- Lewis, M. (2000). *Teaching Collocation: Further Developments in the Lexical Approach*. Hove, England: Language Teaching Publications.
- L'Homme, M.-C. (2019). Combinatoire spécialisée : trois perspectives et des enseignements pour la terminologie. *TTR*, 30(1-2), 215-241.
- Louw, W. E. (1993). Irony in the Text or Insincerity in the Writer? — The Diagnostic Potential of Semantic Prosodies. *Text and Technology: In honour of John Sinclair*, 157-176.
- Louw, B. (2000). Contextual prosodic theory: Bringing semantic prosodies to life. *Words in Context: In honour of John Sinclair*, 48-94.
- McAlpine, J., & Myles, J. (2003). Capturing phraseology in an online dictionary for advanced users of English as a second language: a response to user needs. *System*, 31(1), 71-84.
- Morin, E., & Daille, B. (2006). Comparabilité de corpus et fouille terminologique multilingue. *Traitement Automatique des Langues, ATALA*, 47(1), 113-136.
- Morin, E., & Daille, B. (2012). Compositionnalité et contextes issus de corpus comparables pour la traduction terminologique. *Conférence conjointe JEP-TALN-RECITAL 2012*, 141-154.
- Nattinger, J. R., & De Carrico, J. S. (1992). *Lexical Phrases and Language Teaching (Oxford Applied Linguistics)*. Oxford University Press.
- Nesselhauf, N. (2003). The Use of Collocations by Advanced Learners of English and Some Implications for Teaching. *Applied Linguistics*, 24(2), 223-242.
- Pecman, M. (2007). Approche onomasiologique de la langue scientifique générale. *Revue française de linguistique appliquée*, XII(2), 79-96.
- Pecman, M. (2008). Compilation, formalisation and presentation of bilingual phraseology: Problems and possible solutions. *Phraseology in Foreign Language Learning and Teaching*, 203-222.
- Pecman, M. (2012). Etude lexicographique et discursive des collocations en vue de leur intégration dans une base de données terminologiques. *The Journal of specialised translation (JoSTrans)*, 18, 113-138.
- Peeters, B. (2019). Phraséologie et discours. *French Studies*, 73(4), 669-670.
- Siepmann, D. (2005). Collocation, Colligation and Encoding Dictionaries. Part I: Lexicological Aspects. *International Journal of Lexicography*, 18(4), 409-443.
- Siepmann, D. (2006). Collocations et dictionnaires d'apprentissage onomasiologiques bilingues : questions aux théoriciens et pistes pour l'avenir. *Langue française*, 150(2), 99-118.
- Sinclair, J. M. (1996). The search for units of meaning. *Textus*, 9(1), 76-106.
- Sinclair, J. M. (2004). Meaning in the framework of corpus linguistics. *Lexicographica*, 20, 20-32.
- Tutin, A., & Grossmann, F. (2002). Collocations régulières et irrégulières : esquisse de typologie du phénomène collocatif. *Revue française de linguistique appliquée*, VII(1), 7-25.
- Udovič, B. (2016). Vpliv jezika diplomacije na normo knjižnega jezika. *Slavistična revija: časopis za jezikoslovje in literarne vede*, 64(3), 365-384.
- Willis, D. (1990). *The Lexical Syllabus. A New Approach to Language Teaching*. London and Glasgow: Collins.
- Wray, A. (2002). *Formulaic Language and the Lexicon*. New York : CUP.