Collocations in Portuguese and Slovak: an analysis based on corpora of written and spoken language

JÁN TUPÝ
(Bratislava)

This paper presents an analysis of collocations in Portuguese and Slovak. It aims to spotlight the manifestations of lexical-semantic interference in the sphere of collocations in spontaneous spoken utterances. The analysis is based on the present author's own corpora of spoken language created as a part of his doctoral research (Tupý, 2019a), and it draws on the results of that research as well as on Portuguese and Slovak corpora of written texts, (e.g., the Corpus do Português and the Slovak National Corpus). Portuguese is analysed as a foreign language for Slovak students, with there being a primary focus on its European variant.

KEYWORDS: collocations, interference, Slovak, Portuguese, lexical--semantic level

INTRODUCTION

Collocations are an inherent part of lexis, and, as they are complex units of meaning, their usage in a foreign language often presents considerable problems for students. Both in Lusophone and Slovak linguistics, a number of approaches to the term "collocation"

can be found. In principle, there are two paths of definition. On the one hand, linguists view collocations based on the criterion of frequency, within which they focus on the co-occurrence of two or more structural elements. On the other hand, there are also definitions based on the semantic criterion. For the purposes of this paper, collocations will be defined by taking into account both of these criteria. This definition is based on the works of Slovak and Lusophone linguists, and especially the works of Ďurčo (2007), Jarošová (2007), Majchráková (2005), Mistrík (1993), Ološtiak (2015), Louro (2001), Sanromán (2000), and Tagnin (1998, 2005, 2010).[1]

Based on existing contrastive analyses, it is apparent that collocations in Slovak and their Portuguese equivalents have differences primarily on the morphosyntactic and the lexical-semantic levels.[2] These differences create potential difficulties in formulating utterances in Portuguese and represent potential sources of interference. This paper focuses on an analysis of formulations from a sample of students where there were manifestations of interference on the lexical-semantic level.

1. METHOD

The data for the analysis of manifestations of interference were obtained by means of an experiment carried out on the linguistic production of students of translation and interpreting specializing in Portuguese and another foreign language at the master's level of studies. This choice of respondents was based on the fact that at this level a relatively high command of Portuguese is already expected. The overall sample of the experiment numbered six people,[3] which led

[1] For a more detailed treatment of delimiting collocations, see Tupý (2019a).
[2] Cf. Tupý (2019a) and (2019b).
[3] The size of the sample was also closely linked to the fact that this is the number of students at the Department of Romance Languages at the Faculty of Arts at Comenius University in Bratislava, which is the only institution in Slovakia offering a study programme in Portuguese.

to the need to repeat the process of data acquisition from particular participants several times.

The native language of all the participants was Slovak, and none of them spoke Portuguese as their first foreign language. In addition to Portuguese, the individual students included in the sample had an active command of English, Spanish, French, or German as well as a passive command of Russian or Polish. One student exhibited a relatively strong tendency to use the Brazilian variant of Portuguese on the phonetic level and had a partial penetration of Brazilian features in other levels of the language. These facts were taken into consideration in the analysis. It is necessary to take into account the fact that while the students were bilingual individuals, this bilingualism was asymmetrical and late; furthermore, Portuguese in this context was in the form of an interlanguage.

Since collocations form an inherent part of natural utterances, spontaneous speech was used for obtaining linguistic material. This form was chosen for several reasons. The oral form of communication is more spontaneous than the written form, where students can reconsider collocations or reformulate statement several times and avoid collocations without readers being able to detect such processes. A speech that has been prepared in writing, in turn, gives the speaker the opportunity to verify and seek advice on the collocations before their actual presentation.

In order to eliminate any possibility of preparation by students for the experiment, source texts in Slovak were used that dealt with various topics (the environment, education and the labour market, racial intolerance, health, and social networks) which had not been announced in advance.

Within the analysis of the particular items, there were consultations with a native speaker. For verifying the frequency of Portuguese collocations, their Slovak equivalents, and alternative variants of collocations, the following corpora were used: the Corpus do Português, Corpus de Extractos de Textos Electrónicos MCT/Público (CETEMPúblico), and the Slovak National Corpus.

In the analysis of interference manifestations in the utterances of participants, manifestations of interference were singled out on the lexical-semantic level. The potential pitfalls resulting from the differences between Slovak and Portuguese equivalents serve as

starting points for observing the actual manifestations of interference in the oral communication of the participants.

When analysing the utterances of the participants, the overall acceptability of the used expression was taken into account. Since non-terminological collocations were primarily encountered, it was not the aim to merely evaluate them within the strict opposition of them being 'correct' (OK) or 'wrong' (NotOK). Their level of intelligibility, with an evaluation of expressions being understandable (✓) or incomprehensible (X), and an emphasis on the naturalness of individual expressions in Portuguese were also taken into account. In this categorization, there was a partial reliance on the typology of interferemes according to Veselý (1986). Above all, the criterion of comprehensibility and the general frequency of the used expression were looked at. It is important to point out that the inclusion of the above statements should be considered as indicative. There were four defined groups of expressions:

a) Comprehensible with a high level of naturalness (hereinafter also "✓/OK+")

b) Comprehensible with a lower level of naturalness (hereinafter also "✓/OK~")

c) Comprehensible but erroneous, therefore unnatural (hereinafter also "✓/NotOK")

d) Incomprehensible, erroneous, and therefore unnatural (hereinafter also "X/NotOK")

2. THE ANALYSIS OF THE IDENTIFIED MANIFESTATIONS OF INTERFERENCE

In the analysis of the statements of the participants, there were various manifestations of interference on the lexical-semantic level

which were placed into several subgroups depending on their nature. The findings are based on those from the present author's doctoral research (Tupý, 2019a).

2.1 The transposition of the Slovak component by various Portuguese counterparts

The first (and very numerous) group is represented by expressions where the component in Slovak can be translated into Portuguese in two or more ways if it stands alone, i.e., not in a collocation, e.g., Slovak *volať* 'call'; 'telephone') = Portuguese 1. *chamar*, 2. *telefonar*. However, with regard to the given meaning and the collocates, it is necessary to distinguish between their usage so that the speaker can succeed in transferring the intended meaning.

The need to transpose certain elements of the structure of collocations by different counterparts in equivalent Portuguese lexical sequences stems from the natural fact that the meanings of lexical units across the lexicons of the two languages do not fully correspond with each other, which is closely linked to polysemy.

Some collocates allow for the transposition by various functional units; thus, from the point of view of understanding an utterance and its naturalness, expressions from the category ✓/OK+ were identified that were often fully acceptable and where its components display a high degree of synonymy, e.g., Slovak *učiť sa* 'learn'; 'study' = Portuguese 1. *aprender*, 2. *estudar*. In terms of focusing on interference, it was interesting to observe a strong tendency among participants to prefer the verb aprender over the variant estudar, e.g., in the following statements:

i. *O que eu acho sobre a ideia da cancelação da obrigação de aprender língua inglesa como a primeira língua estrangeira nas escolas primárias...*
ii. *Há muitas razões, por que aprender inglês.*
iii. *Eu acho que é necessário aprender inglês mas a segunda língua devia ser não obrigatória para todos.*

The above tendency presumably stems from associations in the interlingual memory of the students, where the Portuguese *aprender* (and not estudar) functions as the basic counterpart of the Slovak *učiť sa*, as the students primarily associate estudar with the Slovak word *študovať* 'study', which considerably narrows down the meaning of *estudar*.

A manifestation of interference of Slovak can also be found in the following formulation:

iv. *...na* Eslováquia *há muitos professores que não podem trabalhar, porque não há lugares para eles.*

Here the student used the lexical unit lugares, i.e., a direct translation of the Slovak *miesto* 'place'; in Slovak this also means 'job, position'. In this case, the student counted on the context and did not specify the expression by using *de trabalho*. The student preferred the choice of the lexical unit lugar to the alternative expressions *posto* or *vaga*.

In the subcategory ✓/OK ~ there was the following expression:

i. *...cair ao solo...*

The student selected the lexical unit solo instead of the more suitable *chão*, probably associating the latter more closely with the Slovak *podlaha* 'floor'. With regard to this student, however, the influence of the Spanish language cannot be ruled out, cf. the Spanish lexeme *suelo*.

From the subcategory ✓/NotOK, the following formulations deserve a closer look:

i. *...concordo com o que falou* (the speaker makes a correction) *disse a senhora*.

Although Slovak also distinguishes between *hovoriť* 'speak' and *povedať* 'say', the Portuguese counterparts *falar* and *dizer* cannot be used in the same contexts and with the same collocates, e.g., *hovoriť/povedať pravdu* 'to tell the truth' = *dizer a verdade*, not **falar*

a verdade. The above statement (i) points to the unacceptable interchange of these two lexemes. In Slovak, it is appropriate to say *súhlasím s tým, čo pani hovorila*, as well as *súhlasím s tým, čo pani povedala*. The choice of the lexical unit falar can also be supported by Brazilian Portuguese, where it is common in similar contexts. The influence of factors other than the mother tongue has also been observed in the following example:

ii. *Não é o caminho* (searching for the word, a full pause) *o caminho, o melhor caminho que podemos elegir.*

On the basis of the suffix -ir, it can be supposed that the selection of the deformed lexical unit elger instead of the appropriate escolher necessary for the correct collocation *escolher um caminho* is conditioned by Spanish, which uses this component in such contexts.

Similar linguistic behaviour can be observed in the following utterance:

iii. *Não, por exemplo, chamo tanto por teléfono*.

The student used the lexical sequence **chamar por teléfono* instead of the (contextually correct) Portuguese expressions *telefonar*, fazer um telefonema, or *ligar* (the Slovak equivalents of which are *volať* 'call', *telefonovať* 'telephone'). The presumption that there is a significant influence of Spanish is also supported by the fact that the Portuguese word telephone is distorted by the suffix -o and that the accent is in a position typical of the Spanish pronunciation of that term. On the other hand, a few seconds later, the student presents the lexeme *telefonar* in the distorted form *telefonear*, which may demonstrate at least partial knowledge of the alternative of the correct equivalent, albeit with interference from Spanish.

The most critical impact of interference is manifested by the following examples from the subcategory X/NotOK. The first formulation to be stated here stands on the border between the categories ✓/NotOK and X/NotOK:

i. *Isto pode causar que o professor não quer ensinar mais e vai sair da educação*.

In case of excessive effort and thanks to the context, it is possible to partially decipher the intended meaning of the statement as corresponding to the Slovak *odísť zo školstva* 'to leave the educational system', but, on the other hand, a native speaker could interpret it in the sense of 'becoming outraged and giving up good behaviour'. In the word *educação*, the semantic component of education would be activated, which is contrary to the intention of the participant in question. The expression *educação*, with the meaning of 'educational system' can be found, for example, in the Portuguese collocation Ministério da Educação = 'Ministry of Education'; however, in the above statement it is better for it to be replaced by the lexical unit *ensino*, thus creating the collocation *sair do ensino*.

ii. *...e eles foram muito muito muito maus em inglês em apre* (the speaker makes a correction) em educar.

In an attempt to formulate the expression 'to teach English' (cf. Slovak *učiť angličtinu*) = *ensinar inglês*, or *dar aulas de inglês*, one student, under the influence of interference, incorrectly selected the verb *aprender* = Slovak *učiť sa* 'to learn, to study', which the student apparently realized in the middle of the word and replaced with the verb *educar* that, however, cannot be considered as a full-fledged substitute. Not only is it not collocable with inglês, but it is also in sharp contrast with the rest of the statement in terms of style; moreover, in this formulation the recipient perceives it as an unnatural element from a higher style.

Misunderstanding results from using the verb chamar in the following formulation:

iii. *...chamo os* (the speaker makes a correction) *as pessoas que querem trabalhar numa companhia...*

Here again there is a semantic extension common in Slovak, i.e., *volať* 'to call' = *telefonovať* 'to telephone', that is inadmissible in

Portuguese, i.e., *chamar ≠ telefonar*, which has been described above. In this context, moreover, chamar is understood in the sense of 'summoning someone somewhere' and 'inviting someone (e.g., to an interview)', which does not allow for the speaker's intended interpretation of 'to call someone'.

A suitable solution conducive to the elimination of manifestations of interference resulting from the mechanism described above can be seen in a preventively increased emphasis on familiarizing students with the polysemic character of vocabulary. This goal can be more easily achieved by a conscious inclusion of collocations, which the very nature of this issue predestines them for.

2.2 The transposition of a component by a differently motivated unit

In this subcategory, expressions were recorded that show the different motivations of the Slovak and the Portuguese components in the collocation. Apart from it, the given structural elements do not constitute mutual equivalents.

Collocations with components whose transposition is carried out by a differently motivated unit were characterized by a low degree of intelligibility and almost zero naturalness in the participants' statements. Instead, there were erroneous expressions, with the degree of intelligibility being tied exclusively to the context. In addition, the high degree of intelligibility in the previous group was conditioned by the very semantic closeness of the components and the high degree of their transparency; thus, the very nature of the collocations examined in this section, i.e., transposition by a semantically differently motivated unit, predisposes them to the behaviour described above.

Several examples from this category point to the increased demands on the command of active vocabulary needed to create the particular collocation, as evidenced by the statement at the level of ✓/OK~.

i. *...perdendo o seu lugar de* (the speaker makes a correction) *o seu primeiro lugar*.

Based on the context where the student comments on the importance of English and on the use of the preposition de, it can be deduced that the student tried to find the Portuguese equivalent for the collocation *výsostné postavenie* 'unique position' = *lugar de destaque*, but for the sake of fluency the student ultimately turned to the expression o seu primeiro lugar. This collocation is a sufficiently concise (and therefore functional) expression of the idea, although it would sound more natural in Portuguese if the student omitted the numeral *primeiro* and formulated the statement in the following form: *... perdendo o seu lugar*.

From the category ✓/NotOK the following formulations were analysed in more detail:

i. *...mas a segunda língua devia ser* (a full pause) *não obrigatória.*

The structural element nepovinný cannot be transposed into Portuguese literally in the above-mentioned form, i.e., *não obrigatório/a*; it is necessary to use the equivalent *facultativo/a*, or opcional. The fact that the student was looking for a suitable component in the formulation is also evidenced by a full pause being made right before the production of the expression in question.

ii. *Compro muitos iogures e depois eu os tiro no lixo...*

The production of the lexical sequence **tirar no lixo*, which was supposed to express the meaning of the Slovak collocation *(vy)hodiť do smetí/do koša* 'to throw (out)/into the trash/waste-basket' can be explained by a combination of several factors. It is evident that the given participant reached out for a direct translation of the verb hodiť 'to throw', i.e., atirar. However, under the influence of the Spanish language, the student used the lexeme *tirar*. At the same time, it is clear that the student preferred that lexical unit to the variant deitar, which is characteristic of the Portuguese collocation *deitar no lixo*, which does not exist in Spanish.

Here within the subcategory X/NotOK there are also two types of critical situation. The first one is represented by the case where the recipient is unable to decipher the speaker's intended collocation and cannot assign any meaning to the statement:

i. *...mencionou uma cirança que sefreu de tosso negro, não sei se se diz assim*.

The lexical sequence **tosso negro* is the result of a literal transposition of individual components of the Slovak collocation *čierny kašeľ* 'whooping cough'. In Portuguese, however, this corresponds to the equivalent *tosse convulsa* or the monolexical variants *coqueluche* or *pertússis*. From the very statement, it is evident that the participant is not familiar with the given collocation, as the student pointed out. As this terminological unit does not belong to the student's active vocabulary, the student chose a literal translation; but it is not functional.

On the other hand, there was the possibility of an erroneous (i.e., unintended) interpretation by the speaker of the formulated statement:

i. *...alguns bares* (a full pause) *terminaram a vender esta vodka Absolut.*

The formulation used does not describe the intended meaning of 'ending the sale of/to stop selling'; moreover, on the basis of synonymy with acabar, it actually evokes the opposite idea that, after consideration, the bars decided to sell the product in question. Instead of the interference-based *terminar*, the expression *deixar de* should be used.

A detailed look at the following statement reveals several simultaneous manifestations of interference:

i. *...há duas formas de racismo, no mínimo duas formas de racismo. Um é racismo aberto, quando alguém é agressivo e simplesmente não gosta de pessoas doutra cor doutra nacionalidade, mas também há um tipo de racismo fechado, que temos dentro...*

The speaker compares various manifestations of racism, placing them in opposition within two summary designations, namely *racismo aberto* (from *aberto* 'open') and **racismo fechado* (from *fechado* 'closed'). The first one is functional, although the expression *racismo explícito* can also be considered in this situation. While with the expression *racismo aberto* it is possible to observe traces of interference of Slovak, where 'open manifestations of racism' are referred to, in the latter case the participant did not manifest such linguistic behaviour. Analogously, with the help of Slovak, it would be possible to arrive at the appropriate Portuguese expression *racismo latente – latentný/skrytý rasizmus* 'latent/hidden racism'. The collocation used, i.e., *racismo fechado*, is a manifestation of a different mental process where the speaker applied the antonymy *aberto – fechado* in an inappropriate context, which does not indicate interlingual but rather intralingual interference.

CONCLUSION

This analysis identified the occurrence of lexical-semantic interference which to a considerable extent is influenced by the polysemic character of the lexis. As a result of the semantic closeness of the components, the inappropriately used lexical sequences formulated on the basis of the polysemic character of the component manifest a relatively high degree of comprehensibility. However, in the case of interference in collocations with a differently motivated component, the degree of comprehensibility is low and the statements sound rather unnatural. In addition to cross-linguistic interference, there were also manifestations of intralingual interference. Monitoring the occurrence of interference and associated mechanisms that the students experience within their difficulties in formulating collocations provides valuable information for teaching. The results of the research confirm that in order to achieve the formulation of a fully comprehensible and natural statement, it is crucial to systematically draw attention to these phenomena when teaching Portuguese. At the same time, it is necessary to keep in mind the risk of interference from a non-maternal language, e.g., a language typologically closer to

Portuguese. With regard to the findings of broader research of corpora of Slovak collocations and their Portuguese equivalents (Tupý, 2019a), the interaction of factors at the lexical-semantic and morphosyntactic levels presents a potential risk.

REFERENCES

ĎURČO, Peter (2007): *Zásady spracovania slovníka kolokácií slovenského jazyka.* Accessible at http://www.juls.savba.sk/ediela/frazeologicka_terminologia/

JAROŠOVÁ, Alexandra. (2007): Problém vymedzenia kolokácií. In: *Jazykovedný časopis*, 58 (2).

MAJCHRÁKOVÁ, Daniela (2005): *Vyčleňovanie lexikalizovaných spojení pomocou štatistických nástrojov*. Bratislava.

MISTRÍK, Jozef (1993): *Encyklopédia jazykovedy*. Bratislava: Obzor.

OLOŠTIAK, Martin et al. (2015): *Viacslovné pomenovania v slovenčine*. Prešov: Faculty of Arts, University of Prešov.

LOURO, Inês (2001): "*Enxergando" as colocações: para ajudar a vencer o medo de um texto autêntico.* São Paulo: Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo

SANROMÁN, Iriate (2000): *A unidade lexicográfica : palavras, colocações, frasemas, pragmatemas*, Minho: Universidade do Minho.

TAGNIN, Stella (1998): *Convencionalidade e produção de texto: um dicionário de colocações verbais inglês/português português/inglês*. (Livre Docência em Língua Inglesa). São Paulo: Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo.

TAGNIN, Stella (2005): *O jeito que a gente diz: expressões convencionais e idiomáticas*. São Paulo: Disal.

TAGNIN, Stella (2010): Glossário de Linguística de Corpus. In: VANDER Viana; TAGNIN, Stella (Org.). *Corpora no ensino de línguas estrangeiras*. São Paulo: HUB Editorial

TUPÝ, Ján (2019a): *Analýza kolokácií vybraných tematických okruhov so zameraním na interferenciu medzi portugalským a slovenským jazykom*. Bratislava.

TUPÝ, Ján (2019b): Colocações do eslovaco e os seus equivalentes portugueses: divergências e interferência. In. *Romanistica Comeniana*. 2019 (1). pages 11– 21.

VESELÝ, Josef (1986): K otázce překonávání interference. In: *Cizí jazyky ve škole*. 301986/1987 (2).

CORPORA:

O Corpus do Português
Corpus de Extractos de Textos Electrónicos MCT/Público (CETEMPúblico)
Slovak National Corpus

Ján Tupý
Katedra romanistiky, Filozofická fakulta
Univerzita Komenského v Bratislave
Gondova 2, 811 02 Bratislava
jan.tupy@uniba.sk