

Gutiérrez Rubio, Enrique

**A lexicographical study of Czech and Spanish phraseology in spontaneous informal language production**

*Linguistica Brunensia*. 2021, vol. 69, iss. 1, pp. 81-92

ISSN 1803-7410 (print); ISSN 2336-4440 (online)

Stable URL (DOI): <https://doi.org/10.5817/LB2021-1-4>

Stable URL (handle): <https://hdl.handle.net/11222.digilib/144164>

License: [CC BY-NC-ND 4.0 International](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Access Date: 17. 02. 2024

Version: 20220831

Terms of use: Digital Library of the Faculty of Arts, Masaryk University provides access to digitized documents strictly for personal use, unless otherwise specified.

Enrique Gutiérrez Rubio

# A LEXICOGRAPHICAL STUDY OF CZECH AND SPANISH PHRASEOLOGY IN SPONTANEOUS INFORMAL LANGUAGE PRODUCTION

## ABSTRACT

*This paper presents a comparative study that aims at identifying the extent to which the most complete Czech and Spanish phraseological dictionaries are helpful for their users regarding the language employed in spontaneous conversations. Since the oral corpora of these two languages are not equally reliable, the source of the phraseological units (PUs) for this research is conversations extracted from the Czech and the Spanish versions of Big Brother. As a result of the research difficulties faced, only verbal idioms (VIs) are considered. In 90 minutes of conversations among the contestants, 29 Spanish and 34 Czech VIs were documented. These VIs were confronted with Slovník české frazeologie a idiomatiky (SČFI) and Diccionario fraseológico documentado del español actual (DFDEA). According to the data obtained, three main conclusions can be drawn: a) a high degree of difficulty when comparing PUs included in Czech and Spanish phraseological dictionaries, b) both works are relatively reliable tools for users interested in the language employed in spontaneous conversations, at least regarding the verbal idioms, c) in Czech, phraseology has a more strongly colloquial nature than in Spanish.*

## KEYWORDS

*Czech phraseology; Spanish phraseology; spontaneous informal language, lexicography; Big Brother*

## 1 Introduction

In a previous paper (GUTIÉRREZ RUBIO 2018) I presented the results of a pilot study that aimed at identifying the characteristics of Czech phraseological units (PUs) used in spontaneous informal language production, including mainly the frequency

and typology of Czech PUs, but also a lexicographical analysis and a gender-oriented research. In this paper, however, the focus will be exclusively on the lexicographical aspects of phraseology, but instead of the relationships among PUs used in spontaneous conversations in one language being investigated, a cross-linguistic (Czech and Spanish) approach to the research will be proposed. Accordingly, this paper will examine the extent to which the most complete Czech and Spanish phraseological dictionaries are or are not useful tools for the users of these works who need to understand the meaning and the use of PUs uttered in spontaneous conversations (in terms of the number of PUs registered in the dictionaries and the register labels included in them). Moreover, conclusions regarding general differences in the nature of the Czech and Spanish phraseological systems will be drawn.

## 2 Oral corpora and *Big Brother* as a source of authentic spontaneous informal language

The main topic of the research presented in this paper is the use of Czech and Spanish phraseology in spontaneous informal language production. Although the existing oral corpora of these two languages are reasonably large, they are not equally reliable. While the Czech National Corpus' ORAL2013<sup>1</sup> contains around 2,800,000 words, the latest version of the oral corpus of the present-day Spanish language GREA oral 3.2<sup>2</sup> gathers almost nine million words. Interestingly, it is the shorter corpus, i.e. the Czech ORAL2013, which is the more consistent of the two, since the 1,297 Czech speakers from all regions of the Czech Republic recorded for this specific purpose were not aware that they were being taped. On the contrary, the GREA oral corpus consists of all kinds of oral linguistic material from more than 1,600 sources including, among others, political speeches, phone conversations, messages on answering machines, and informal dialogues.<sup>3</sup> Because of the lack of consistency of the Spanish oral corpus, for the comparative research presented in this paper I have opted for linguistic data recorded under very similar circumstances. One of the possibilities at our disposal was to use conversations from radio or TV programmes. However, this sort of data is far from fulfilling the main "colloquialiser" features proposed by Briz Gómez (2010), i.e. a relationship of equality, experience-based correlation, an everyday interaction frame, and quotidian topics. On the contrary, it is not questionable that these features are ideally fulfilled by the contestants of the reality show *Big Brother* because of the shared situation of

1 See <<https://wiki.korpus.cz/doku.php/en:cnk:oral2013>>.

2 See <<http://corpus.rae.es/creanet.html>>.

3 See <<http://www.rae.es/rekursos/banco-de-datos/crea-oral>>.

the “captives” in the house. Moreover, they are all recorded under exactly the same circumstances.

So, when taking into account the pros and cons of *Big Brother* as a source of oral linguistic material, it has to be admitted, on the one hand, that it is a reality show and that, logically, it merely “simulates situations”. In consequence, we cannot be dealing here with 100 % *authentic* linguistic expressions. On the other, Dovey states that, as a material practice, this kind of simulation “[...] produces real knowledge about real things in the real world and has real effects upon real lives” (DOVEY 2004, 233). Moreover, Penadés Martínez (2004, 2227) confirms that, at least in the interviews and conversations that she analysed, the degree of “formality” tended to decrease clearly in the last minutes of the recordings, which lasted approximately 45-60 minutes. In accordance with all this, it seems to be acceptable to proceed from the hypothesis that, after being recorded and broadcast live 24/7 for several weeks, the participants could probably behave linguistically almost as if they were not being recorded, i.e. they would be making use of spontaneous informal language production. In fact, a preliminary analysis confronting the data extracted from *Big Brother* with that recollected by Penadés Martínez (see GUTIÉRREZ RUBIO 2019) sufficiently shows that the PUs (concretely *verbal idioms*) excerpted from the reality show are labelled “colloquial” 2.1 times more frequently and labelled “vulgar” 28 times more frequently than those obtained from Penadés Martínez’ corpus (for more information on this topic in Spanish, see GUTIÉRREZ RUBIO 2021). According to this methodological approach, the linguistic material for this research was excerpted from *Gran Hermano*, the Spanish version of the original TV franchise, and *VyVolení*, a Czech variation of *Big Brother* based on a Hungarian licence.

### 3 Methodology of the analysis

#### 3.1 *Gran Hermano*

For this study 90 minutes of the reality contest *Gran Hermano*, broadcast in Spain by the private television station Tele 5, were analysed. These 90 minutes of programme belonged to the 59<sup>th</sup> and 68<sup>th</sup> days of the 16th season, which was broadcast live 24/7 via online streaming between September and December 2015. From the original eighteen contestants, only ten remained in the house in that moment, specifically four women and six men. Their average age was 22.9 years and their socio-economic status ranged from middle- to upper-middle-class, as illustrated by the fact that three of the *Gran Hermano* contestants (Han, Sofía, and Marta) were pursuing higher education at the time they participated in the show.

The fragments were randomly recorded by myself from the 24/7 online streaming channel and subsequently analysed, i.e. I personally viewed the videos several times and excerpted all the PUs from the conversations that were monitored.

### 3.2 VyVolení

The reality contest *VyVolení* was broadcast by the Czech private television station TV Prima. The fragments that were analysed for this study belonged to the 26<sup>th</sup> and 47<sup>th</sup> days<sup>4</sup> of the fourth season of this reality show, which was broadcast live 24/7 via online streaming between August and November 2013. Fourteen contestants (eight women and six men) took part in the conversations included in these fragments. Their average age was 30.7 and their socio-economic status ranged from lower-middle- to middle-class.<sup>5</sup>

In view of the fact that I am not a native speaker of Czech, it was my Czech (native) students who, after having been instructed for this purpose, excerpted the PUs from the fragments. In total, every student examined one hour of *VyVolení*. In order to increase the reliability of the data that was obtained, a second student watched the video again and reviewed the data presented by the first student. Finally, a third Czech student checked the data again.<sup>6</sup> Of course, the whole procedure was conducted under my personal supervision.

### 3.3 Czech and Spanish phraseological dictionaries

In the case of the Czech language, there can be no doubt that the *best* phraseological dictionary is *Slovník české frazeologie a idiomatiky (SČFI)* by Čermák et al. (2009). In fact, it is certainly the only reliable dictionary of Czech phraseology ever published. The more than 3,500 pages of its four main volumes incorporate approximately 35,000 PUs. Unquestionably, this monumental work belongs among the largest and most complete dictionaries of phraseology in the world. As mentioned above, it is structured into four main volumes, elaborated by a total of twenty compilers through approximately four decades. The first three books were originally published as individual works: 1. *Přirovnání* ('Similes', first published in 1983), 2. *Výrazy neslovesné* ('Non-verbal phrasemes', 1988), and 3. *Výrazy slovesné* ('Verb-based phrasemes', 1994). Fifteen years later, in 2009, these three books were re-edited, revisited, and completed with new phraseological material, specifically with a fourth volume entitled *Výrazy větné* ('Propositional phrasemes'). Finally, in 2016 a very useful tool (although it included no new PUs) entitled *Onomaziologický slovník* ('Onomasiological dictionary') was added to the four main volumes.

This dictionary is not just an impressive achievement because of its enormous number of entries. Additionally, the quantity and quality of the information that accompanies the PUs gathered here is remarkable: together with the core data,

4     Respectively retrieved from the following links: <[https://www.youtube.com/watch?v=ii\\_yLE6q4To](https://www.youtube.com/watch?v=ii_yLE6q4To)> and <<https://www.youtube.com/watch?v=RSd4MtIoSuk&spfreload=10>> (last access: 17/12/16).

5     For instance, none of the Czech contestants had a college degree or was pursuing higher education at that time.

6     I would like to express my deep gratitude to the students of the course "Frazeologie a obraz světa" ('Phraseology and Picture of the World') and to Kateřina Lamaczová, who checked the final data.

such as the meaning and use of every expression, a series of further pieces of information is included, such as variations in form, stylistic and morphological labels, intonation, or multilingual equivalents (in English, French, German, and Russian). Nevertheless, the dictionary shows some problems regarding the systematicity of its elaboration. The cause can primarily be attributed to the fact that the four main volumes were elaborated by different teams of specialists working in different periods and under different conditions. It has to be admitted that in the final edition published in 2009, an effort was made to present a standardised format in all the tomes. However, each volume displays specific features.

Moreover, and this question is central for this research, in *SČFI* there are two kinds of entries – “hesla běžná/plná” (‘regular/full entries’), which provide the user with all the information mentioned above, and “hesla okrajová/redukovaná” (‘peripheral/reduced entries’), in which only the very basic information can be found, i.e., meaning, use, and variations of the PU in question. Interestingly, every volume has its own proportions of these two kinds of entries. So, whilst in the first tome (*Similes*) almost all the entries are *regular*, in the third and fourth volumes 62 % and 69 % of the total are *reduced*; in other words, most of the PUs registered in these two volumes lack stylistic labels. However, the authors claim that the most frequent PUs have full entries. If this is true, in our corpus the percentages of reduced entries should be markedly less than this.

Unlike *SČFI*, *Diccionario fraseológico documentado del español actual: locuciones y modismos españoles (DFDEA)* by Seco et al. (2004) is not the only *reliable* dictionary of Spanish phraseology. Nevertheless, it has been chosen for this study since it is undeniably the most complete, extensive, and up-to-date one among the Spanish phraseological dictionaries published so far. It is true that some specialists, mainly Penadés Martínez (2015a), have criticised a certain lack of systematicity in some of its aspects, but even so, it must be admitted that it has been elaborated on a systematic lexicographical basis. Moreover, this work is up-to-date and documented, which means that all the PUs are illustrated with examples excerpted from (European) Spanish texts dated between 1955 and 2004. On the other hand, *DFDEA*, with about 16,000 entries in 1,084 pages, is not as extensive as *SČFI*. Moreover, *DFDEA* collects less information about the entries in comparison with *SČFI* – only their meaning, use, stylistic labels, and, as already mentioned, at least one short textual example. In spite of this, for the purposes of this study *DFDEA* seems to be a more reliable source of information than *SČFI*, given that all its entries include the key information needed with regard to the aim of this research, i.e. the stylistic label; in other words, whether the PU in question belongs to the colloquial register or not. In general, it can be stated that *DFDEA* is a much more systematic work than *SČFI*, probably because it was composed by just three compilers in a relatively short period of time.

### 3.4 Research difficulties

One of the most common struggles in cross-linguistic research is the lack of consistency of the data that is being compared. In this sense, it must be assumed that the Czech and Spanish dictionaries chosen as the sources of data do not gather and classify PUs according to exactly the same methodological criteria. In fact, the Czech and the Spanish phraseological traditions have historically followed different paths and, to some extent, nowadays study slightly different phenomena. Despite these dissimilar traditions and the noticeable lexicographical differences mentioned in section 3.3, both dictionaries declare that they pursue the same aim – to collect PUs that are common in the contemporary language. Moreover, both agree on what could be called the central notion of phraseology, whatever it may be, but certainly one very close to the general English concept of *idiom*. Accordingly, a PU such as the Czech *Dělat si blázny/blázna z někoho* or Spanish *Tomarle el pelo a alguien* (in both cases meaning roughly ‘make a fool of someone’) will always be gathered in any extensive phraseological dictionary. On the other hand, this is not necessarily true for those expressions located on the periphery of the phraseological territory. Probably the most salient methodological divergence in this sense, although not the only one, is the inclusion in *SČFI* of proverbs and their exclusion from *DFDEA*. It seems that the Czech phraseological tradition accepts proverbs, whilst the Spanish specialists mostly agree in distinguishing between phraseology and paremiology, devoted exclusively to the study of proverbs, at least in the field of phraseography.

Unfortunately, the different categorisation and taxonomy of the PUs is just one of the research difficulties faced in this study. The second problem, unlike the first one, is not caused by the fact that the approach used is a cross-linguistic research study, but by the methodological approach adopted in this study. As already mentioned in sections 3.1 and 3.2, different coders excerpted the PUs from the fragments of *Gran Hermano* and *VyVolení*, the Czech videos being combed by Czech students and the Spanish ones by myself. Despite the fact that the students were thoroughly instructed to recognise the PUs and that three individuals checked every fragment, the data clearly shows that many PUs were overlooked by the Czech students.

Besides these two main research difficulties, two more minor divergences could influence the results of the research. Firstly, although *Big Brother* and *VyVolení* share the same format, a few discrepancies can be found regarding these two reality shows, mainly in terms of the sociolinguistic characteristics of the participants, with the Czech housemates being noticeably older (almost eight years on average) and being of lower socio-economic status (see in sections 3.1 and 3.2) than the Spanish contestants. Second, the very notion of the stylistic labels (including *colloquial*) used in the dictionaries could be added to the lexicographical discrepancies mentioned above. In this sense, Seco et al. (2004, xxix) propose a clear definition of *colloquial* as a label used for PUs typically used in informal conversations or relaxed

situations (“Uso propio de la conversación informal o de una situación distendida”).<sup>7</sup> On the other hand, Čermák et al. (2009) admit that the proposed stylistic labels have vague boundaries and that, consequently, they have merely an “orientational” character. Maybe because of this, the authors do not propose any clear definition of the label “kolokviální”; they just note that most PUs have a colloquial nature, which is a marker for the prototypically spoken informal language (“Většina frazémů má tedy pak povahu kolokviální, což je označení pro prototypicky mluvený neformální jazyk”, ČERMÁK et al. 2009/iv, 13).

With the intention of minimizing the effects of the above-mentioned research difficulties, this study focuses exclusively on verbal idioms (VIs), and this for several reasons: a) unlike other types of PU (proverbs, routine formulae, collocations, and some types of idioms), there is a general consensus regarding the recognition of VIs as PUs; b) VIs are one of the most salient and clearly recognisable types of PUs, which should drastically reduce the consequences of having used different coders; c) from all the types of PUs VIs are the only object of study of our main reference works in Spanish (PENADÉS MARTÍNEZ 2004; 2012; 2015b).

## 4 Results of the analysis

During the 90 minutes of conversations, 29 VIs (types) were documented in Spanish and 34 VIs in Czech. However, and because of the methodology used, this data tells us nothing of scientific value about the frequency distribution of PU in both languages, which is in no case the goal of this study, although it does show that, at least in terms of VIs, we are dealing here with substantially similar data. One of the basic goals of this study is to present the percentage of expressions registered and not registered in the dictionaries. Whilst in *SČFI* 21 VIs are included (61.8 % of all the Czech VIs), in *DFDEA* 19 VIs are registered (65.5 % of all the Spanish PUs). Consequently, both dictionaries seem to be almost equally efficient tools for the users of these works who need to understand the meaning and the use of PUs uttered in spontaneous language production, at least in terms of VIs.

Another aim of this research was to investigate the stylistic labels associated with the excerpted PUs. The first interesting piece of data is that eight out of the 21 registered Czech VIs (38.1 %) lack stylistic labels, i.e. belong to what are termed the “reduced” entries. Although in some volumes the majority of the entries were reduced (62 % in the case of the third tome, the one including the VIs), this, in comparison, relatively low number of “reduced” entries documented in this study is not fully unexpected, since, as mentioned in section 3.3, the most frequent PUs should

7 Moreover, rude/vulgar PUs are also colloquial and can be defined as bad expressions that should not be used in the presence of respectable people (“Expresión malsonante o del mal gusto que no debe emitirse ante personas de cierto respeto”, SECO et al. 2004, xxxii).



have a full entry in the dictionary. Moreover, the percentages regarding those VIs that do have register labels are presented in Table 1.

**Table 1.** Stylistic registers for verbal idioms in *SČFI* and *DFDEA* (only PUs with labels)

Type	Colloquial	Rude or vulgar	Neutral	Others	Total
<i>SČFI</i> (Czech)	8 (61.5 %)	2 (15.4 %)	3 (23.1 %)	0 (0 %)	13 (100 %)
<i>DFDEA</i> (Spanish)	6 (31.6 %)	2 (10.5 %)	11 (57.9 %)	0 (0 %)	19 (100 %)

According to the data obtained only for the VIs with stylistic labels, 76.9 % of the Czech VIs are colloquial (61.5 % plus 15.4 %, since all the rude or vulgar PUs are also considered colloquial). Although it is true that the conversations took place in an informal context, these numbers are unexpectedly high, especially when confronted with the data obtained for Spanish in very similar communicative situations – a mere 42.1 % of colloquial VIs (31.6 % plus 10.5 %) against 57.9 % of neutral PUs.

The explanation for these divergences could be found in a series of reasons. As already mentioned, the results are almost certainly affected by the fact that two different methodologies were used for obtaining the data. However, a more probable key factor could be the lexicographic methodology the compilers made use of. In this sense, Čermák et al. (2009/I, 9) state that in Czech most of the phraseology has a strongly spoken nature (“Většina frazeologie má v češtině výrazně mluvenou povahu”). According to this, it is plausible to think that, when it came to labelling the PUs, Čermák and his team members tended to use the “colloquial” label more often than the Spanish compilers. One illustrative example, although it was not documented in the 90 minutes of conversations from *VyVolení* or *Gran Hermano*, could be the PUs used in these languages to wish someone a hearty appetite, i.e. *dobrou chuť* in Czech and *que aproveche* in Spanish. So, whilst this PU is registered in *DFDEA* with the label “neutral”, in *SČFI* it is “colloquial/neutral”, which, according to the methodology of this research, would be considered colloquial, as this label is written in the first place (before “neutral”).

Nonetheless, after the Czech VIs labelled “colloquial” are confronted with their Spanish equivalents proposed by Dubský in his Czech-Spanish dictionary (*Velký česko-španělský slovník (VČŠS)*, 1996), all the Czech informal PUs excerpted from *VyVolení* have at least one equivalent in VČŠS labelled “colloquial” in *DFDEA*. In fact, almost all the VIs excerpted from *VyVolení* are undoubtedly informal, such as *mocť bejt někomu ukradeněj* (‘can take a running jump/a long jump off a short pier, couldn’t give a damn about it/him’, according to ČERMÁK et al. 2009/iii, 840) and its Spanish version (*me importa*) *un blede* (SECO et al. 2004, 189) or *dělat si z něčeho/někoho srán-*

du ('make fun of sth/so., take the mickey/piss out of so.', ČERMÁK et al. 2009/iii, 341) and its Spanish equivalent *tomar a coña*, which is even labelled "vulgar" by Seco et al. (2004, 315). Accordingly, from the ten colloquial and vulgar VIs documented in *VyVolení* and labelled in *SČFI* only one PU could arouse doubts about its informal nature: *nebýt ve své kůži* ('be out of sorts', ČERMÁK et al. 2009/iii, 334). In this sense, *be out of sorts* is not labelled in *The Oxford Dictionary of Idioms*, which means that it is considered "neutral" by its author (SIEFRING 2004, 270). On the other hand, in Spanish, Dubský proposes two equivalents to *nebýt ve své kůži* that, at the same time, are registered in *DFDEA* – the first one, (*no estar*) *a sus anchas*, is a neutral PU (SECO et al. 2004, 141); on the contrary, the second one, (*tener*) *mal cuerpo*, is labelled "colloquial" (SECO et al. 2004, 351).

Therefore, examining the raw data and some controversial examples, such as *dobrou chuť*, Čermák et al. (2009), unlike Seco et al. (2004), could seem to use a different lexicographical methodology regarding colloquial and neutral labels, according to which they would magnify the presence of informality in the Czech phraseology. However, after providing a deeper interpretation, it is possible to infer that Čermák et al. are probably right when they treat spoken and colloquial registers in Czech as being almost equal. This could be explained by the historical development of this language, especially during the 17th and 18th centuries, because of which there would be a larger gap between formal and informal registers than in other European languages with a more continuous linguistic development, such as Spanish or English.

A third variable for explaining the differences may well be the socio-economic status of the contestants, since the Spanish speakers are almost eight years younger on average and could be considered to belong to a higher class than their Czech counterparts. In my opinion, this could only explain the observed divergences to some extent, although further research with a younger and higher-class group of Czech speakers should be performed in order to shed light on this issue. However, when I asked my Czech students for feedback about their analysis of *VyVolení*, one of the most common comments concerned the rude vocabulary used by the contestants. It is true that, unlike in Spanish, in the Czech media it is extremely uncommon to hear swearwords and vulgar statements (very probably caused by the already-mentioned bigger gap between formal and informal registers in Czech). The fact that the frequency of "vulgar" VIs is five points higher in Czech (15.4 %)⁸ than in Spanish (10.5 %)⁹ is a decidedly unexpected piece of data that speaks for the plausible influence of the sociolinguistic factor. According to this, the data shows an extraordinarily informal register during the conversations that took place in

8 Two Czech vulgar VIs were documented: *lézt někomu do prdele* ('lick so.'s arse', ČERMÁK et al. 2009/III, 615) and *posrat to* ('bugger it up, make a cock up of sth', ČERMÁK et al. 2009/III, 589).

9 Two Spanish vulgar VIs were documented: *no tener ni puta idea* (roughly 'not to have a fucking clue') and *tener cojones (para)* (roughly 'to have the bollocks (to)').

*VyVolení*, especially in terms of the recurrent use of vulgar expressions, a register that other Czech oral corpora, in which the speakers will make efforts to use a more formal register, would probably not be likely to document. This fact would speak for the relevance of using *VyVolení* as a source of highly spontaneous informal language production in Czech. However, to know the degree to which this is so, a further study would be necessary, especially contrasting this data with that obtained from the Czech National Corpus' ORAL2013.

## 5 Conclusion

Despite all the methodological difficulties faced in the research from which the data presented in this paper was obtained, some conclusions can be drawn.

First, if every comparative or cross-linguistic research study is *per se* problematic, the degree of difficulty increases still further when dealing with phraseology, probably because this linguistic discipline does not yet have clear boundaries and different (national) traditions understand the object of its study from (at least slightly) different perspectives. In addition, these distinct theoretical approaches are inevitably reflected in the methodological criteria employed by the specialists to compile PUs in their dictionaries. Moreover, these discrepancies can also be observed in the stylistic labels associated with the PUs registered in them. This seems to be the case of the two most complete dictionaries of Czech and Spanish phraseology.

Second, according to the data obtained, it can be stated that both *DFDEA* and *SČFI* are relatively reliable tools for users interested in the language employed in spontaneous conversations, at least regarding verbal idioms. However, the Spanish dictionary seems to be slightly more useful, not just because it registers almost 4 % more VIs than the *SČFI* does (65.5 % vs. 61.8 % of all the documented VIs), but especially since only 62 % of all the VIs excerpted from *VyVolení* include a stylistic label in *SČFI*.

Third, almost 77 % of all the VIs excerpted from the Czech conversations and labelled in *SČFI* are “colloquial” or “vulgar”. This percentage can be considered unexpectedly high, at least when compared with the 42 % of colloquial and rude entries in its Spanish counterpart. This deviation of 35 points seems not to be primarily caused by the methodological inaccuracies of this research – although it must be admitted that this fact could have influenced the data obtained – or by the dissimilar linguistic and lexicographical approaches used by the Czech and Spanish compilers, but by the fact that in Czech phraseology has a more strongly colloquial nature than in Spanish.

Fourth, the uneven socio-economic status of the speakers, with the Czech contestants being older and from a lower class, could explain the higher frequency of PUs labelled “rude” or “vulgar” in Czech (15.4 %) than in Spanish (10.5 %).

Finally, an extraordinarily informal register has been evidenced during the conversations that took place in *VyVolení*, especially in terms of the frequent use of vulgar expressions, a register that other Czech oral corpora would probably not be likely to display. According to this, *VyVolení* would be a better source of spontaneous informal language production than the Czech National Corpus' ORAL2013. However, to explain the degree to which this is so, a further study would be necessary.

## REFERENCES

- BRIZ GÓMEZ, Antonio. 2010. Lo coloquial y lo formal, el eje de la variedad lingüística. In: CASTAÑER MARTÍN, Rosa María – LAGÜENS GRACIA, Vicente, coords. *De moneda nunca usada: Estudios dedicados a José M<sup>a</sup> Enguita Utrilla*. Zaragoza: Insituto Fernando El Católico, CSIC, 125–133.
- ČERMÁK, František et al. 2009. *Slovník česke frazeologie a idiomatiky (I. Příkladování, II. Výrazy neslovesné, III. Výrazy slovesné, IV. Výrazy větné) (SČFI)*. Praha: Leda.
- CORPAS PASTOR, Gloria. 1996. *Manual de fraseología española*. Madrid: Gredos.
- DOVEY, Jon. 2004. It's only a game show: Big Brother and the theatre of spontaneity. In: MATHIJS, Ernest – JONES, Janet, eds. *Big Brother International: Formats, Critics and Publics*. London/New York: Wallflower Press, 232–248.
- DUBSKÝ, Josef. 1996. *Velký česko-španělský slovník/Gran diccionario checo-español*. Praha: Leda.
- GUTIÉRREZ RUBIO, Enrique. 2018. Czech phraseology in spontaneous informal language production. A pilot study. In: GUTIÉRREZ RUBIO, Enrique et al., eds. *Contributions to the 21st Annual Scientific Conference of the Association of Slavists (Polyslav)*. Wiesbaden: Harrassowitz, 76–83.
- GUTIÉRREZ RUBIO, Enrique. 2019. Typologies of phraseological units and their frequency in the reality show *Gran Hermano*. *Phrasis. Studi fraseologici e paremiologici* 3, 246–252.
- Gutiérrez Rubio, Enrique. 2021. *Fraseología española en el discurso oral*. Valencia: Tirant lo Blanch.
- PENADÉS MARTÍNEZ, Inmaculada. 2004. ¿Caracterizan las locuciones el registro coloquial? In: VILLAYANDRE LLAMAZARES, Milka, ed. *Actas del V Congreso de Lingüística General, León, 5-8 de marzo de 2002, III*. Madrid: Arco Libros, 2225–2235.
- PENADÉS MARTÍNEZ, Inmaculada. 2015a. Implicaciones de la frecuencia de uso de las locuciones en la elaboración de un diccionario. *Estudios de Lingüística Universidad de Alicante (ELUA)* 29, 253–277.
- PENADÉS MARTÍNEZ, Inmaculada. 2015b. Las locuciones verbales en el habla de Madrid (distrito de Salamanca). In: CESTERO MANCERA, Ana M. – MOLINA MARTOS, Isabel – PAREDES GARCÍA, Florentino, eds. *Patrones sociolingüísticos de Madrid*. Bern: Peter Lang, 251–286.
- SECO, Manuel et al. 2006. *Diccionario fraseológico documentado del español actual: locuciones y modismos españoles (DFDEA)*. Madrid: Aguilar.
- SIEFRING, Judith, ed. 2004. *Oxford Dictionary of Idioms*. Oxford: Oxford University Press.



Enrique Gutiérrez Rubio

*A lexicographical study of Czech and Spanish phraseology in spontaneous informal language production*

Acknowledgement: I would like to gratefully acknowledge the financial support for this research provided by the project “Nové trendy ve výzkumu literatury, lingvistiky a kultury v románských zemích” (IGA\_FF\_2020\_023).

*Enrique Gutiérrez Rubio*

Department of Romance Languages

Faculty of Arts, Palacký University Olomouc

Křížkovského 10, 771 80 Olomouc

Czech Republic

enrique.gutierrez@upol.cz



This work can be used in accordance with the Creative Commons BY-NC-ND 4.0 International license terms and conditions (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>). This does not apply to works or elements (such as image or photographs) that are used in the work under a contractual license or exception or limitation to relevant rights.