Typologies of phraseological units and their frequencyin the reality show Gran Hermano

Enrique Gutiérrez Rubio

UNIVERSITY PALACKÝ OLOMOUC (CZECH REPUBLIC)

Abstract: This paper presents a study that aims at identifying the frequency and typology of Spanish phraseological units (PUs) used in spontaneous informal language production. For this kind of study, it is necessary to have access to spontaneous, free-spoken conversations. The source of the PUs for this research is formed by conversations extracted from *Gran Hermano*, the Spanish version of *Big Brother*. In this study data obtained from the analysis of almost two hours of spontaneous conversations extracted from *Gran Hermano* 16 (broadcast live 24/7 via on-line streaming) will be presented and discussed. Moreover, it is intended to open up for discussion whether this reality show can be a better source of spontaneous informal conversations than the existing Spanish oral corpora[1].

→**Keywords**: spontaneous informal language, frequency of Spanish phraseology, typology of Spanish phraseology, *Gran Hermano*

1. Introduction[2]

→ This paper aims to shed some light on the frequency and typology of Spanish phraseology as it is used in spontaneous informal language production. For this kind of analysis, it is necessary to have access to spontaneous, free-spoken conversations. Since such linguistic material cannot be found in radio or TV programmes, where the speakers tend to use the standard language, the source of the PUs analysed in this research is *Gran Hermano*, the Spanish version of the TV show *Big Brother*. A secondary aim is to test the hypothesis that this reality show could be a better source of spontaneous conversations than the existing Spanish oral corpora (this topic will be discussed in more detail in Section 2.3).

→ In this sense, although some of the oral corpora of Spanish are unquestionably large, they are not sufficiently reliable for the purposes of this research study. The best-known Spanish oral corpus is the oral version of the Corpus de Referencia del Español Actual (CREA) composed by Real Academia de la Lengua Española. The

---

1 See http://corpus.rae.es/creanet.html (consulted 13.11.2019).

most recent edition of this present-day Spanish language corpus, CREA oral 3.2[3], gathers almost nine million words. Nevertheless, it consists of all kinds of oral linguistic material from more than 1600 sources including, among others, political speeches, phone conversations, messages on answering machines, and informal dialogues[4]. Because of the lack of consistency of this Spanish oral corpus, for the research study presented in this paper, I have opted for linguistic data with more consistency. As mentioned above, conversations from radio or TV programmes are one of the potential sources available to us. However, this sort of oral data is far from fulfilling the main *colloquialiser* features proposed by Briz Gómez (2010: 126), i.e. a relationship of equality, experience-based correlation, an everyday interaction frame, and quotidian topics. On the contrary, it is reasonable to admit that these features are ideally fulfilled by the contestants of the reality show *Gran Hermano* because of the shared situation of the *captives* in the house. Another advantage of this methodological approach is that, unlike many oral corpora, such as CREA oral, all the speakers are recorded under identical circumstances.

→ In this sense, when the pros and cons of *Gran Hermano* as a source of oral linguistic material are considered, it must be admitted that it is a reality show and that, logically, it merely *simulates situations*. In consequence, the conversations cannot be evaluated as 100% *authentic* linguistic material. Nevertheless, there are authors, mainly researchers working from a sociological perspective, who claim that these shows produce effects that are more real than one might

think. In this vein, Dovey (2004: 233) 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 ». However, these opinions are not enough to confirm the validity of these conversations as a reliable source of spontaneous informal language production, a topic that will be discussed in more detail in Section 2.3.

→ In this study, the data obtained from *Gran Hermano* will be compared with the phraseological units (PUs) extracted by Penadés Martínez from the PRESEEA ALCALÁ oral corpus[5]. This corpus consists of eighteen conversations and sixteen interviews in Alcalá de Henares (Madrid). The recordings lasted nearly 45-60 minutes and the corpus consists of approximately twenty-seven hours' worth of material and 322000 words (see Penadés Martínez, 2004). Despite all the efforts made, she admits that this oral corpus cannot be considered colloquial, but that it varies between formal and semiformal registers. Interestingly, Penadés Martínez (2004) confirms that the degree of *formality* showed a clear decreasing tendency in the last minutes of the recordings; occasionally, it decreased to the extent that some speakers admitted that they even forgot that they were being recorded at all, although during the interviews and talks the recorder was always in sight. In the case of *Gran Hermano,* the conversations that were examined took place on the 67th and 76th days of the programme, i.e. at a time when the participants had been monitored and broadcast live 24/7 for more than 1500 hours. Consequently, if after just a few dozen minutes the speakers involved in PRESEEA ALCALÁ oral corpus tended to use a more colloquial register, it can be proposed that the contestants of

---

3 See http://corpus.rae.es/creanet.html (consulted 13.11.2019).
4 See http://www.rae.es/recursos/banco-de-datos/crea-oral (consulted 13.11.2019).

5 See http://preseea.linguas.net/Corpus.aspx (consulted 13.11.2019).

Gran Hermano could behave linguistically almost as if they were not being recorded, at least after having lived together for weeks in conditions that were highly conducive to *colloquialisation*. In other words, they could be making use of spontaneous informal language production.

2. Methodology of the analysis

→ For this pilot study, 112 minutes of the reality contest *Gran Hermano*, broadcast in Spain by the private television station Tele 5, were analysed. These almost two hours of programming belonged to the 67th (85 minutes) and 76th (27 minutes) days of the sixteenth season, which was broadcast live 24/7 via online streaming between September and December 2015. From the original eighteen contestants, four women (Marta, Niedziela, Marina, and Sofía) and six men (Carlos, Aritz, Suso, Han, Ricky, and Vera) remained in the house at that moment. Their ages varied between 19 and 28 and their average age was 22.9 years; 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. While eight participants were originally Spanish, two were born abroad. Han comes from China, but he moved (together with his mother) to Armilla (Granada) when he was 12 years old[6]. He is not a native speaker of Spanish, but in 2015, when he participated in the TV competition, he was very fluent in Spanish and studied Electrical Engineering at the Universidad de Granada[7]. Moreover, Daniel Vera was

born in Argentina and, unlike Han, he is a native speaker of Spanish–at the time of the show he had already been living in Spain (specifically Gran Canaria) for sixteen years and had a strong Canarian Spanish accent.

→ The fragments were randomly recorded from the 24/7 online streaming channel and subsequently analysed, i.e. the videos were viewed several times and all the idioms, routine formulae, and proverbs (see section 2.2 for further discussion on these three concepts) from the conversations that were monitored were excerpted.

2.1 Frequency distribution

One of the aims of this research on phraseology as it is used in natural communicative situations is to reveal the frequency of the use of PUs. In total, 536 occurrences and 259 types were documented in these 112 minutes of linguistic material. Logically, some PUs appear more than once in the conversations. To be precise, 173 PUs are documented once, 38 PUs twice, seventeen PUs three times, eight PUs four times, six PUs five times, and so on. The PU *Es que* is the most frequent expression, since it is documented forty times, followed by the PU *O sea*, with twenty-nine occurrences. Besides, since 9565 words are documented during the 112 minutes that were analysed, on average one different type is used every 36.9 words and one occurrence every 17.8 words. Since no previous studies had investigated this particular question, it is not possible to evaluate whether these frequencies are high or low. However, in the near future a similar analysis will be performed regarding the frequency of phraseology in two kinds of language production – in spontaneous formal language production (oral corpus of the present-day Spanish

---

6 See https://www.lecturas.com/actualidad/desvelamos-el-duro-pasado-de-han_17606 (consultado 13.11.2019).
7 See http://vertele.eldiario.es/verteletv/actualidad/concursantes-Gran-Hermano_0_1720327956.html (consulted 13.11.2019).

language CREA oral 3.2[8]) and in spontaneous informal language production (Val.Es.Co. corpus[9], a compilation of 43 secretly recorded conversations[10]).

2.2 Typology

→ One of the main methodological difficulties faced by phraseologists is that it is rather exceptional to find two specialists who use the same typology and the same terminology. After the evaluation of many existing taxonomies, three basic categories of PUs are proposed for this research.

1. *Idioms* (*locuciones* in Spanish) such as *Tomarle el pelo a alguien* [make a fool of someone] are frequently considered the most salient type of PUs. Although the boundaries of this group are not absolutely defined, in the sense proposed in this research, an idiom would be a unit « [...] that is fixed and semantically opaque or metaphorical, or, traditionally, "not the sum of its parts", for example, *kick the bucket* or *spill the beans* » (Moon, 1998: 4).

2. *Routine formulae* (*fórmulas oracionales* in Spanish) are a heterogeneous group of multiword expressions frequently used for pragmatic purposes–independent speech acts expressing a complete thought–such as Spanish *¿Sabes lo que te quiero decir?* [You know what I mean?].

3. *Proverbs* (*paremias* in Spanish) are independent maxims with a deontic function such as

Type	Occurrences	%	Types	%
Idioms	370	(69.0%)	164	(63.3%)
Routine formulae	165	(30.8%)	94	(36.3%)
Proverbs	1	(0.2%)	1	(0.4%)
Total	536	(100%)	259	(100%)

Table 1: Typology of Spanish PUs

Spanish *A quien madruga, dios le ayuda* [the early bird catches the worm].

The results of the typological analysis of the PUs registered in the recordings of *Gran Hermano* are presented in Table 1.

→ As can be observed in Table 1, the presence of the type *idioms* is clearly dominant among the PUs that were documented. Nevertheless, the most relevant data seems to be the very low frequency of *proverbs*. Moreover, the relatively high number of occurrences in the case of *routine formulae*, with around a third of all instances, seems to indicate that a frequent use of this type of PU is a distinctive feature of conversations in spontaneous informal language production.

2.3 Gran Hermano as a reliable source of spontaneous informal language

→ One of the aims of this research study was to identify the extent to which this reality show can or cannot be considered a source of *natural* spontaneous informal language production. To shed some light on this issue, the data extracted from *Gran Hermano* was confronted with the PUs documented by Penadés Martínez (2004) in her own study based on the PRESEEA oral corpus. Since, in her paper, this Spanish linguist does not provide us with all the documented PUs, but

---

8   See http://corpus.rae.es/creanet.html (consulted 13.11.2019).
9   See http://www.valesco.es/?q=corpus (consulted 13.11.2019).
10   See http://www.valesco.es/ (consulted 13.11.2019).

only with a list of verbal phrases (*luciones verbales* in Spanish), only this specific subtype of idioms will be evaluated. From all the verbal phrases excerpted from Penadés Martínez' work, 147 types (with 496 occurrences) are registered in *Diccionario fraseológico documentado del español actual: locuciones ymodismos españoles (DFDEA)* by Seco *et al.* (2004), the most complete, extensive, and up-to-date of the Spanish phraseological dictionaries published so far. On the other hand, from the forty-four verbal phrases extracted from our recordings of *Gran Hermano,* thirty-one (with fifty-three occurrences) are registered in *DFDEA.* In Table 2, the stylistic labels proposed for the excerpted PUs by Seco *et al.* (2004) in their dictionary are presented.

belled *colloquial* 2.1 times more frequently and labelled *vulgar* 28 times more frequently than those obtained from Penadés Martínez' study. This data clearly shows that the speakers of *Gran Hermano* use a much more colloquial register than those who participated in the conversations and interviews analysed by Penadés Martínez. Accordingly, it can be stated that *Gran Hermano* is a better source of spontaneous conversations than (at least some of) the existing Spanish oral corpora.

Corpus		Colloquial (vulgar)	Neutral
Penadés Martínez	Types	43.6% (0.7%)	56.4%
	Occurrences	18.1% (0.2%)	81.9%
Gran Hermano	Types	48.4% (9.7%)	51.6%
	Occurrences	37.7% (5.6%)	62.3%

Table 2: Stylistic registers in DFDEA

→ So, the PUs extracted from *Gran Hermano* are labelled *colloquial* and especially *vulgar*[11] more frequently than those PUs from the PRESEEA oral corpus (in this sense, it is important to note that all the vulgar PUs are also considered colloquial). As a result of the important differences regarding the size of both corpora in question, for the purposes of this study it seems more appropriate to compare occurrences than types. According to this, in *Gran Hermano* PUs are la-

3. Conclusion

→ The study presented in this paper is just a first attempt to investigate some key aspects of Spanish phraseology as it is used in spontaneous, free-spoken conversations. Specifically, an analysis of 112 minutes of conversations among ten contestants (four women and six men) of the sixteenth season (September-December 2015) of the reality show *Gran Hermano*, the Spanish version of *Big Brother*.

11 In Penadés Martínez (2004) work, just one vulgar PU out of 496 occurrences was documented: *tocar los huevos.*

→ Although we are dealing here with a pilot study, some promising conclusions regarding the *real use* of phraseology in spontaneous informal language production can be drawn:

I. A huge difference in terms of the frequency of use among the PUs that were documented has been revealed. Whilst most of them (67% of the total) are documented just once, the routine formulae *Es que* (an expression that could be translated into English as *the thing is...*, *well...*, or even *it's just that...*) is employed forty times in less than two hours of conversations.

II. Moreover, on average one occurrence of a PU is used every 17.8 words. However, because of the lack of previous studies investigating this issue, whether this frequency is high or low cannot be evaluated.

III. The typological analysis presented in this paper shows that idioms are the most frequent type of PUs, followed by routine formulae, a type with a relatively high number of occurrences, probably because they are a distinctive feature of the phraseology used in spontaneous informal language production. Proverbs, with just one occurrence out of 536 PUs, are clearly rare in this register.

IV. In *Gran Hermano* PUs are labelled colloquial 2.1 times more frequently and labelled vulgar 28 times more frequently than those PUs included in the oral corpus analysed by Penadés Martínez. According to this data, it can be stated that this reality show is a reliable source of spontaneous informal conversations.

→ Although far more data must be analysed in the future, the results obtained from this pilot study provide some relevant clues about the nature of Spanish PUs used in spontaneous informal language production.

References

BRIZ GÓMEZ, Antonio (2010), "Lo coloquial y lo formal, el eje de la variedad lingüística", in Castañer Martín, RosaM. / Lagüéns Gracia, Vicente (ed.), *De moneda nunca usada: Estudios dedicados a José Mª Enguita Utrilla*, Zaragoza, Institución Fernando el Católico, 125-133.

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, 232-248.

MOON, Rosamund (1998), *Fixed Expressions and Idioms in English. A Corpus-Based Approach*, Oxford, Clarendon Press.

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.

SECO, Manuel *et al.* (2006), *Diccionario fraseológico documentado del español actual: locuciones y modismos españoles (= DFDEA)*, Madrid, Aguilar.

Corpus

Corpus CREA RAE, http://corpus.rae.es/creanet.html (consulted 13.11.2019).

Corpus ORAL RAE, http://www.rae.es/recursos/banco-de-datos/crea-oral (consulted 13.11.2019).

Corpus PRESEEA, http://preseea.linguas.net/Corpus.aspx (consulted 13.11.2019).

Corpus VALESCO, http://www.valesco.es/ (consulted 13.11.2019).

Han Gran Hermano *El Diario*, http://vertele.eldiario.es/verteletv/actualidad/concursantes-Gran-Hermano_0_1720327956.html (consulted 13.11.2019).

Han Gran Hermano *Lecturas*, https://www.lecturas.com/actualidad/desvelamos-el-duro-pasado-de-han_17606 (consulted 13.11.2019).

Bio-bibliographic profile

Enrique Gutiérrez Rubio completed his PhD in Diachronic Czech Linguistics in 2007, at Universidad Complutense de Madrid (Spain). He has been Assistant Professor at the Department of Romance Languages (Faculty of Arts) of Palacký University Olomouc (Czech Republic) since 2008. He has published four books and over thirty peer-reviewed journal articles and book chapters, mainly on Spanish and Czech morphology and phraseology.

E-mail: enrique.gutierrez@upol.cz