***L. Abalo-Dieste, J. Pérez-Guerra***

**PASSIVISATION AND RELATIVISATION  
AS COLLOQUIALISATION STRATEGIES   
IN PRESENT-DAY ENGLISH: A CORPUS-BASED STUDY**[[1]](#footnote-1)

**Abstract.** This corpus-based study aims to determine whether colloquialisation is a process affecting written English or a general tendency of the English language that can also be attested in speech. To that end, this study investigates two syntactic clausal processes evincing variation in twenty-first century English due to a process of colloquialisation: passivisation and relativisation. The first phenomenon is explored through the overall productivity of passive predicates, and of *be*- and *get*-passive constructions. The second phenomenon is investigated by analysing choices between, on the one hand, *who* and *whom*, and *which* versus *that*/zero relativisers. The data are collected from the Spoken BNC2014 and the BNC1994DS corpora, as well as from the F-LOB and the BE06 corpora. The findings support the hypothesis that passivisation and relativisation are determinants of colloquialisation both in spoken and in written English.

**Keywords.** colloquialisation, written, spoken, passive, relative.

**1. Introduction**

Traditional distinctions between speech and writing have been blurred due to the demands of present-day communication. The growing proximity between writing and speech is called ‘colloquialisation’ and constitutes the topic of this study. Recent linguistic research on colloquialisation has evinced a growing degree of orality and simplicity techniques in written English in all varieties and text types. Specifically, this paper investigates (potentially) representative linguistic features of colloquialisation in spoken Present-Day British English (BrE) with the aim of determining whether colloquialisation is a phenomenon specific of written English or a tendency affecting the English language as a whole, which, in consequence, is also shown in spoken English.

This study is organised as follows. Section 2 summarises previous work on colloquialisation. Section 3 describes the purpose of the current corpus-based research, followed by a description of the methods used to collect, retrieve and analyse the data. Section 4 reports the empirical results and discusses the main findings. Finally, Section 5 contains the summary and the concluding remarks, with a focus on the contribution of this chapter to the study of colloquialisation.

**2. Colloquialisation: An overview**

This section summarises previous studies on colloquialisation. Colloquialisation is defined by [Iosef 2013: v] as “the growing influence of speech on written language”. In studies like [Collins, Yao 2013: 480], this process of language change is associated with the influence exerted by speech on the written medium, speech being regarded as a combination of features typical of face-to-face conversation and characteristics of written language. [Iosef 2013: 133] argues that this process of growing orality does not only involve informality in writing, but also the incorporation of “more direct speech” in written texts. [Collins, Yao 2013: 480] contend that colloquialisation spreads not only to written genres but to spoken language as well, which justifies this investigation.

Linguistic features associated with colloquialisation have been investigated through corpus analyses. In [Collins, Yao 2013] the effect of colloquialisation in non-native varieties of English is assessed by exploring data from ICE corpora through features such as contractions, semi-modals, *get*-passives and the contraction *let’s*. Another investigation measuring the significance of colloquialisation in English is [Iosef 2013]. This author analyses contractions, quotative *like*, and phrasal verbs with *up* and *out*. [Collins 2008] and [Levin 2013] carry out descriptive studies of the distribution of specific linguistic features, such as modal auxiliaries and semi-modals. [Baker 2017: 237–238]. compares BrE and American English (AmE) by paying attention to spelling, vocabulary and grammar differences, and takes colloquialisation as an explanatory theory of the results obtained. Regarding colloquialisation, he investigated constructions with empty subjects *it/there*, past participles (as representative of passive constructions), modal auxiliaries and relative pronouns. [Leech et al. 2009: 19] investigate the Brown family of corpora and check the role of prescriptive grammatical approaches to language description. Their study reveals a decrease in the frequency of modals and passives, and an increase of progressives and contractions.

**3. The case study**

This section deals with the purpose, methodology and data retrieval. In Section 3.1 I describe the aim of this research. Section 3.2 tackles the research design and method alongside the criteria for the selection of the linguistic features. In Section 3.3 I present the specific search queries and their limitations.

***3.1. Aims and scope***

A corpus-based analysis has been conducted to explore linguistic features associated with the phenomenon of colloquialisation in the recently released spoken part of the British National Corpus (BNC) 2014, alongside in its predecessor the British National Corpus. The diachronic study of spoken English is compared with the results on written English provided by [Leech et al. 2009] and [Baker 2017: 175] in an attempt to determine whether colloquialisation is a process that also affects spoken BrE or not.

***3.2. Data and methodology***

This investigation is based on data from the demographically-sampled (DS) part of the original BNC spoken part (BNC1994DS) and the spoken section of the BNC2014 [Love et al. 2017: 324]. The data have been retrieved through the Lancaster University’s CQPweb platform [Hardie 2012].

Drawing on [Leech et al. 2009] and [Baker 2017: 176], relativisation and passivisation as investigated here as potential strategies of colloquialisation in spoken BrE. As regards passivisation, I have explored *be-* and *get-*passives. Although *be*-passives outnumber *get-*passives, [Leech et al. 2009: 244] showed that (informal) *get-*passives were gaining ground, particularly in text types closely related to conversation, whereas *be*-passives have become less frequent. Since *be*-passives are considered formal passive constructions which convey objectivity in writing, their declining tendency strengthens the hypothesis of growing colloquiality in BrE in favour of active voice and informal passive variants such as *get*-passives. From this perspective, this paper investigates not only the frequency of *be*- and *get-*passives in spoken English, but also the overall decline of passive voice as a feature not evincing colloquialisation. Relativisation phenomena have been approached in the literature by exploring the distribution of *wh-, that-* and zero-relative clauses. [Leech et al. 2009: 288] take the declining trend of *wh-*relative clauses to be a sign against colloquialisation. Specifically, the frequency of the *who* relativiser decreases alongside other *wh-*forms in BrE, and increases in AmE. [Baker 2017: 167] also reports the decline of *wh-*relative clauses, particularly *which-*clauses, in written BrE, in favour of their more colloquial and growing *that-*relative counterparts. [Leech et al. 2009: 231] found zero-relative clauses to be increasing and, thus, competing with *that-* and *which*-relativisers, while [Baker 2017: 168] reports decreasing frequencies of this relativisation type since 1961. In this paper I investigate the variation of the different relativisation options in an attempt to discern their influence on colloquialisation in speech.

The spoken data have been retrieved from the two 1994 and 2014 BNC corpora. Previous research results have been taken as the benchmark for written BrE. When neither [Baker 2017: 151–175] nor [Leech et al. 2009] provides frequencies for a specific feature, the feature’s frequency is retrieved from the F-LOB (BrE 1991) and the BE06 (BrE 2006) corpora of the Brown family.

The retrieval of the relevant constructions was carried through the CQPweb platform provided by Lancaster University. The F-LOB, BE06, BNC1994DS and Spoken BNC2014 corpora have been annotated with different CLAWS tag-sets, which required the design of specific queries for each corpus. Regarding relative clauses, queries were kept as similar as possible in order to maximise comparability (e.g., **[\_AT (\_{A})?\_{N} (\_{N})? which (I|you|he|she| it|we|they)]** for *which-, that-* and zero-relative clauses, which retrieves instances of an article followed by an optional adjective, followed by a noun, followed by an optional noun, followed by the relativiser *which,* followed by a personal nominative pronoun, likely to function as the subject of the relative clause, as in (1):

(1) apart from the different business culture which you’ve mentioned there’s also ... (Spoken BNC2014, SP2Y 2707)

A comparable query for *who* and *whom* required a complex query through CQP syntax. This query was designed to avoid prepositions before *whom* and attributive complements after *who*: **[!(pos= ‘II|IO|IW’)] ‘who’ (‘I|you|he|she|it|we|they’) [!(pos= ‘VB.\*’)]**. It retrieved instances of *who* not preceded by a preposition and followed by a personal pronoun not followed by any *be* verbal form. Passive search strings were designed to grant comparison. The query **[pos= ‘VB.\*’] [pos= ‘XX|R.\*’]? [(pos= ‘VVN’) & !(word= ‘used|married’)]** detected forms of the verb *be*, followed by an optional negative particle (*not/n’t*)or an adverb, followed by a past participle, excluding *used* and *married.* The query for *get-*passives is practically identical, with a ‘**get|gets|got**’ slot.

**4. Results and discussion**

This section is devoted to the analysis of relativisation and passivisation. Passivisation has been explored through the distribution of *be-* and *get-*passive constructions in the corpora. The raw and normalised (‘Fpw’, per 1,000,000 words) frequencies of the constructions are given in Table 1.

*Table 1*. Passivisation

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Spoken | | | |  | Written | | | |
|  | BNC1994DS | | Spoken BNC2014 | |  | F-LOB | | BE06 | |
|  | Raw | Fpw | Raw | Fpw |  | Raw | Fpw | Raw | Fpw |
| *Be*-passives | 30737 | 2565.02 | 22921 | 2006.63 |  | 11165 | 9768.51 | 9754 | 8503.20 |
| *Get*-passives | 2594 | 216.47 | 4442 | 388.87 |  | 79 | 69.11 | 80 | 69.74 |
| Active voice | 2253002 | 188014.6 | 2538495 | 222234 |  | 162912 | 142535.4 | 165809 | 144546.6 |

In light of the normalised frequencies in Table 1, we hypothesised a number of tendencies: (i) greater frequency of passive constructions in written English (see Figure 2), (ii) decline in the use of the passive voice from the twentieth to the twenty-first century, and (iii) greater frequency of *get-*passives in spoken English (see Figure 1).

|  |  |
| --- | --- |
|  |  |
| *Fig. 1*. *Be*- and *get*-passives | *Fig. 2*. Passivisation frequencies |

In order to test these hypotheses, I compared the frequencies of active versus passive sentences, and *be-* versus *get-*passives, in twentieth and twenty-first written and spoken BrE. The statistical tests reveal highly significant differences between passive and active constructions in written versus spoken twentieth-century English (F-LOB vs. BNC1994DS; χ2=22729.82; *p*<.00001) and in twenty-first-century English (BE06 vs. Spoken BNC2014; χ2=25231.94; *p*<.00001). These results confirm the preference for passive voice in writing, which reflects the association of this construction to formal language. Regarding the productivity of passive voice over time, the difference between active and passive constructions was tested in twentieth- versus twenty-first-century written (χ2=113.53; *p*<.00001) and spoken (χ2=1499.53; *p*<.00001) English. The statistically significant decrease of this feature indicates a movement away from complex constructions and towards colloquialisation. A closer look at the two variants reveals a highly significant difference between *be-* and *get-*passive constructions in written versus spoken twentieth-century English (F-LOB vs. BNC1994DS; χ2=747.63; *p*<.00001) and twenty-first-century English (BE06 vs. Spoken BNC2014; χ2=1610.76; *p*<.00001). Although the prototypical *be-*passive is still the most frequent structure, *get-*passives seem to be gaining ground as informal alternatives in speech. These findings evince that passivisation is a determinant of colloquialisation in English, not only in writing but also in speech.

Regarding relativisation, my analysis focuses on two choices between relativisation devices which differ in style: who- vs. whom-relativisers, and which- vs. that-/zero-relativisers. Attention was paid to the distribution of who- and whom-relativisers in linguistic that allow for alternation with no semantic consequences. Table 2 and, more visually, Figure 3 provide the frequencies, which reflect the greater frequency of whom in written English, and the decrease in the use of whom from twentieth- to twenty-first-century English.

*Table 2*. Relativisation: *who/whom*

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Spoken | | | |  | Written | | | |
|  | BNC1994DS | | Spoken BNC2014 | |  | F-LOB | | BE06 | |
|  | Raw | Fpw | Raw | Fpw |  | Raw | Fpw | Raw | Fpw |
| *Who* | 450 | 37.553 | 654 | 57.255 |  | 10 | 8.749 | 19 | 16.564 |
| *Whom* | 50 | 4.173 | 4 | 0.35 |  | 35 | 30.622 | 10 | 8.718 |

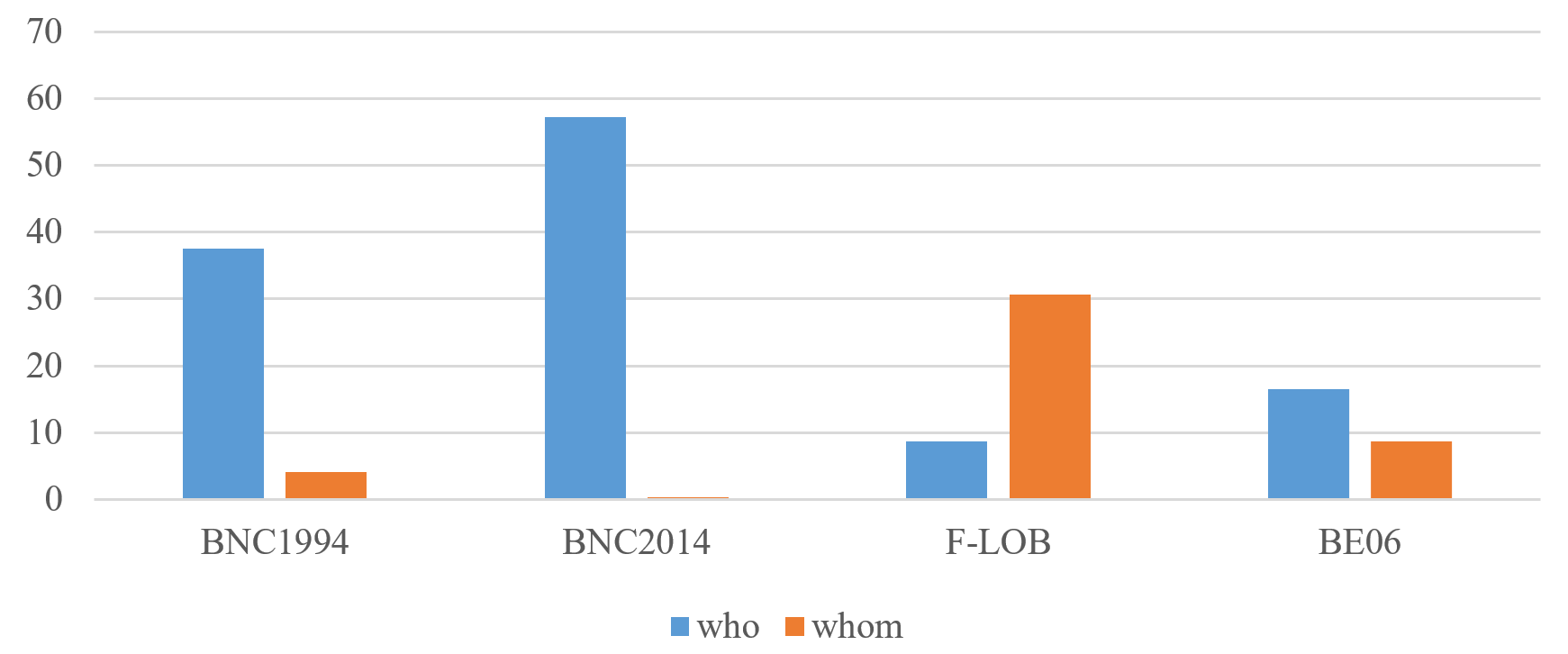


Fig. 3. Relativisation: who/whom

The tests reveal significant differences for written (χ2=13.87; p<.00019) and spoken (Fisher; p<.00001) English. Also, differences between the use of who and whom in spoken versus written twentieth- (χ2=114.07; p<.00001) and twenty-first-century English (χ2=159.66; p<.00001) proved significant. Who is more productive than whom in both speech and writing, and this trend becomes more salient over time. The data also reflect a greater frequency of whom in written English, thus evincing a connection between whom and textual formality, and between who and speech. As suggested in the literature, these findings reveal that who vs. whom variation can be taken as a colloquialisation strategy in written and spoken English. The second stylistic choice affecting relativisation contrasted which- with that- and zero-relative clauses. Following the competition between which- and that-clauses suggested in the literature, I combined the frequencies of these two relativisers, both considered informal alternatives to which-clauses. Table 3 and Figure 4 reveal, first, a higher frequency of that- and zero-relative clauses in spoken English and, second, an increase of these relativisers from twentieth- to twenty-first-century spoken English and decrease in written English.

*Table 3*. Relativisation: *which*, *that* and zero

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Spoken | | | |  | Written | | | |
|  | BNC1994DS | | Spoken BNC2014 | |  | F-LOB | | BE06 | |
|  | Raw | Fpw | Raw | Fpw |  | Raw | Fpw | Raw | Fpw |
| *Which* | 756 | 63.08 | 212 | 18.56 |  | 41 | 35.87 | 17 | 14.82 |
| *That* | 5869 | 489.77 | 4050 | 354.56 |  | 228 | 199.48 | 179 | 156.04 |
| Zero | 25276 | 2109.3 | 30937 | 2708.39 |  | 1064 | 930.91 | 1059 | 923.2 |

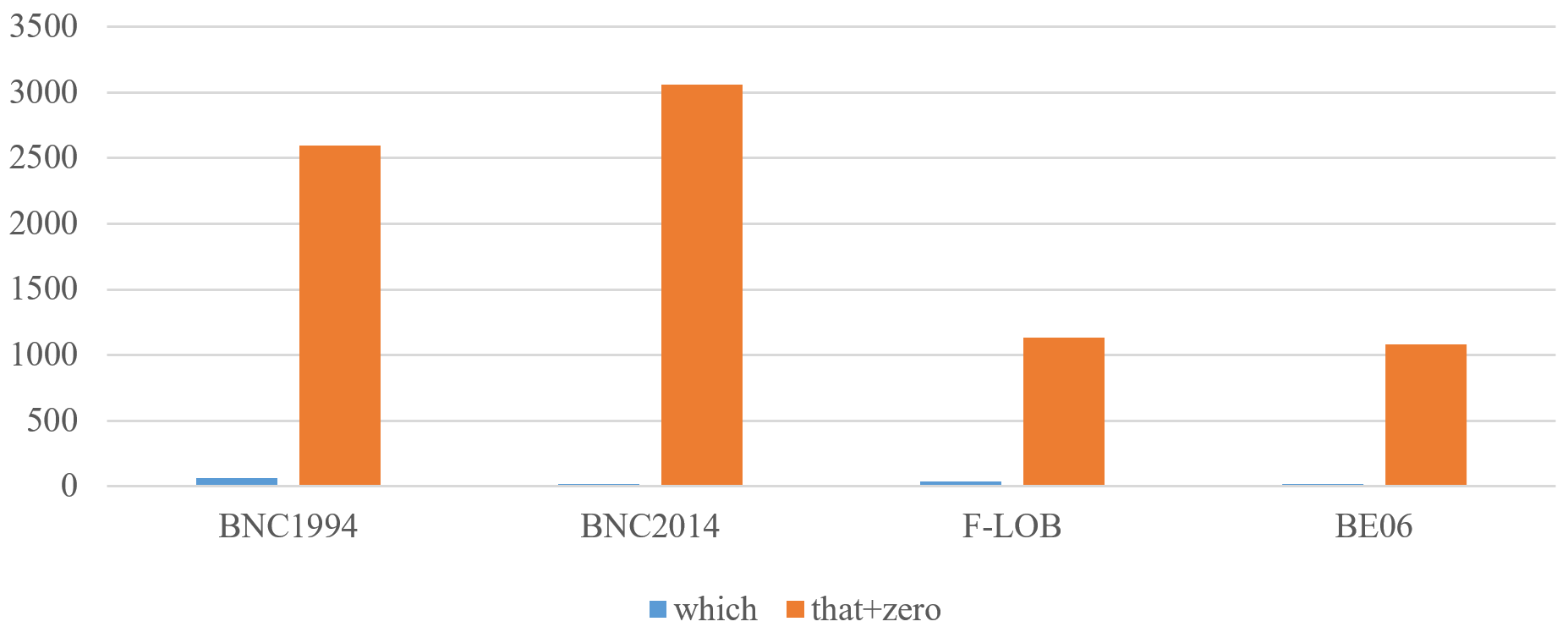


Fig. 4. Relativisation: which, that and zero

The analysis of relativisers showed that the difference between *which*- and *that*-/zero-relative clauses is not significant in twentieth-century written English (χ2=2.72; *p*=.098833). By contrast, this choice is statistically significant in twentieth- and twenty-first-century spoken English (χ2=10.99; *p*<.000918). This suggests that the alternation *which* versus *that/*zero is a relevant to colloquialisation in spoken English in both centuries, while in written English it is only significant as a colloquialisation strategy in the twenty-first century.

**4. Conclusion**

This study has explored two potential colloquialisation strategies in spoken BrE: passivisation and relativisation with the purpose of determining whether the same trends are observed or not in written and in spoken English. In light of data retrieved from twentieth- and twenty-first-century spoken and written English, this study corroborated the connection of passive voice with textual formality and its decline in present time, which contributes to the colloquialisation process. Secondly, the choice between the relativisation options (*who* vs. *whom*, *wh*- vs. *that*-/zero relativisers) proved to be a marker of colloquialisation in writing and speech. These results not only support the colloquialisation hypothesis in spoken BrE, but also suggest that this phenomenon started first in spoken English.

**References**

1. *Baker, P.* (2017), American and British English: Divided by a Common Language? Cambridge: Cambridge University Press.

2. *Collins, P.* (2008), The English Modals and Semi-modals: Regional and Stylistic Variation. In Nevalainen, T., Taavitsainen, I., Pahta, P., Korhonen, M. (eds.), The Dynamics of Linguistic Variation: Corpus Evidence on English Past and Present. Amsterdam: John Benjamins, pp. 129–145.

3. *Collins, P., Yao, X.* (2013), Colloquial features in World Englishes. International Journal of Corpus Linguistics, Vol. 18 (4), pp. 479-505.

4. *Hardie, A.* (2012), CQPweb: Combining Power, Flexibility and Usability in a Corpus Analysis Yool. International Journal of Corpus Linguistics, Vol. 17 (3), pp. 380–409.

5. *Iosef, M.* (2013), Signs of Colloquialization: Three Corpus-based Case Studies. Oslo: Universitetet i Oslo, Master’s Thesis.

6. *Leech, G., Hundt, M., Mair, Ch., Smith, N.* (2009), Change in Contemporary English: A Grammatical Study. Cambridge: Cambridge University Press.

7. *Levin, M.* (2013), The Progressive Verb in Modern American English. In Aarts, B., Close, J., Leech, G., Wallis, S. (eds.), The Verb Phrase in English: Investigating Recent Language Change with Corpora. Cambridge: Cambridge University Press, pp. 187–216.

8. *Love, R., Dembry, C., Hardie, A., Brezina, V., McEnery, T.* (2017), The Spoken BNC2014: Designing and Building a Spoken Corpus of Everyday Conversations. International Journal of Corpus Linguistics, Vol. 22 (3), pp. 319–344.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Laura Abalo-Dieste / Javier Pérez-Guerra**

Universidade de Vigo (Spain).

***E-mail: laura2abalo@gmail.com / jperez@uvigo.es***

1. We would like to acknowledge and thank the Spanish State Research Agency and the European Regional Development Fund (grant no. FFI2016-77018-P), and the Regional Government of Galicia (grant no. ED431C 2017/50) for their generous financial support. [↑](#footnote-ref-1)