The *British Sentimental Novel Corpus* (*BSNC*) and the ROC-DDC alternation at the level of the individual

Tamara Bouso, University of the Balearic Islands

Pablo Ruano San Segundo, University of Extremadura

Abstract

This article provides a description of the *British Sentimental Novel Corpus* (*BSNC*) and a case study that explores, at the level of the individual, the relation between the Reaction Object Construction (ROC) and the Direct Discourse Construction (DDC). The *BSNC* is a large-scale specialised corpus that comprises full novels of eleven canonical authors across three generations from 19th century British fiction. It aims at studying language as both a social and a cognitive phenomenon and, in line with a recent trend in historical sociolinguistics, at exploring the interaction between individual and aggregate levels (see Fonteyn 2017; Hilpert 2020; Petré et al. 2019). The first part describes the methodological principles that underlie the design and compilation of the *BSNC*. In the second part, we present a new case study that aims to determine whether our previous aggregate findings also hold at the individual level. The results serve to confirm our hypothesis: first, individual changes in the ROC and the DDC run in parallel across almost the entire 19th century, correlating most significantly between 1851 and 1860. Second, the aggregate-level division of labour between these two functionally similar constructions turned out to be a feature of all authors in the *BSNC*. Last, the ROC-DDC alternation has been attested in an important proportion of the *BSNC* novels, with only a relatively small group of texts using solely the older and less extravagant variant (i.e. the DDC). This suggests that the alternation as such represents a cognitive reality for these individual writers across their lifespan.

Keywords: *British Sentimental Novel Corpus* (*BSNC*), Reaction Object Construction (ROC), Direct Discourse Construction (DDC), ROC-DDC alternation, individual and aggregate levels, 19th century fiction

1. Introduction

The purpose of this paper is to contribute to the analysis of English language variation and change through the lens of the British sentimental novel, a subtype of novel from the 19th century (Carter & McRae 1996; Head 2006; Ousby 1988) that shows a particular emphasis on 'emotional response' and 'feeling' (Rowland 2008: 193). This emphasis on 'feeling' has not only been documented in the form of brief comments by literary critics (e.g. Baldick 2001; Hunt et al. 1806), but it has also been demonstrated empirically via a sentiment analysis (Piper & Jean So

Bouso, Tamara, and Pablo Ruano San Segundo. 2021. "The *British Sentimental Novel Corpus* (*BSNC*) and the ROC-DDC alternation at the level of the individual." *Nordic Journal of English Studies* 20(1): 215–257.

2015), i.e. 'an area of computational linguistics that analyses people's sentiments and opinions regarding different objects or topics' (Kim & Klinger 2019: 2). Although the British sentimental novel can be traced back to the mid-18th century—Henry Brooke's *Fool of Quality* (1765-1770) is considered to be one of the earliest sentimental novels—, it would not be until the late 18th century when this 'emotionally extravagant novel' (Baldick 2001: 234) became popular. Sentimentality would become, in fact, a defining trait of English fiction in the 19th century, in both the Romantic (1798-1836) and the Victorian (1837-1880) periods. To some extent, it has even been argued in the literature that sentimentality was one of the aspects that contributed to the rise of the novel as a mass genre (Pykett 1994). For instance, it is well known that authors such as Dickens depended upon sensation, sentiment, and melodrama to attract and satisfy new unsophisticated readers (Brook 1970: 143–144). To be more precise, the growing reading public (i.e. women, children, or the emerging working class) were fond of sentimentality, and this fitted nicely in the serialised mode of publication: plots, relationships, characters were all marked by trifling circumstances and sentiments, which kept the audience engaged from one instalment to the next (Altick 1957). In this paper, we present a corpus of this type of novel, which we call the *British Sentimental Novel Corpus* (henceforth *BSNC*), the idea of which arose as a result of a convergence of interests between the two authors of the paper, who in June 2019 at the University of Extremadura decided to test the tight relation between the novel of sensibility and the emergence and development of the so-called Reaction Object Construction (henceforth ROC; e.g. *She smiled disbelief*) (Bouso 2021; Levin 1993: 98).

The *British Sentimental Novel Corpus* (*BSNC*) that we present here is different from other less homogeneous corpora of 19th century fiction such as the *Corpus of Late Modern English Texts, version 3.0* (CLMET3.0, 1710-1920, comprising 34,386,225 words: see De Smet, Diller & Tyrkkö 2013) and *A Representative Corpus of Historical English Registers, version 3.2* (*ARCHER-3.2*, 1600-1999, with c. 3.3 million words: see Denison & Yáñez-Bouza 2012). The former, in its narrative fiction section, covers sentimental novels but also children's books, autobiographies, satires, philosophical novels, diaries, journals, early science fiction and even short stories. The latter (i.e. *ARCHER-3.2*) is a multi-genre corpus of British and American English covering the

period 1600-1999. Its section devoted to 19th century British fiction samples 20 novels written by 21 authors and amounts to c. 90,000 words (see Yáñez-Bouza 2011a, 2011b).

The *BSNC* aims to supplement these multi-genre 19th century corpora and other specialised historical corpora on narrative fiction such as the *19th Century British Fiction Corpus* (*HUM19UK*, 1800-1899, with c. 13 million words: see Walker et al. 2019), the *19th Century Reference Corpus* (c. 4.5 million words) hosted in the *CLiC* (*Corpus Linguistics in Cheshire*) web app (Mahlberg et al. 2016), or even the *Corpus of English Novels* (*CEN*, 1881-1922, totalling 26,227,428 words: see De Smet 2008), which draws on fiction written by both American and British writers at the turn of the 20th century. As we shall see in section 2, the *BSNC* differs from these other similar historical corpora in composition, scope and size. Finally, it should be mentioned here that like *CLMET3.0* (*CLMET* = 9,818,326 words, and *CLMETEV* = 14,970,622 words: see De Smet 2005a, 2006), the *BSNC* should not be understood at this stage as a 'fixed body of texts in the same way conventional corpora of English are; the corpus can be extended or reduced at wish' (De Smet 2005b: 70). We aim therefore to keep on improving the *BSNC* in the future just in the same way that other analogous corpora have been considerably expanded and enhanced over the years (see, for instance, *CLMET3.1*, the tagged version of *CLMET3.0*: De Smet et al. 2015). In what follows, we will discuss the make-up of the corpus and the selection criteria (section 2). We will also illustrate the potential of the corpus by (i) surveying some current research for which the corpus has already been used, and (ii) presenting a new case study on the ROC-DDC alternation (e.g. *She smiled disbelief* and *She smiled, 'I don't believe you'*) from an individual-level perspective (section 3); finally, we will discuss some of its highlights and limitations (section 4).

2. Corpus Make-Up and Selection Criteria

The *BSNC* has been entirely compiled on the basis of texts drawn from *Project Gutenberg*. It is made up of 114 complete novels written by 11 canonical Romantic and Victorian novelists. The novels are listed (in chronological order) in the Appendix. There we have additionally included information about the author (name, surname and lifespan) and the amount of text each novel contributes. These 114 novels total more

than 21 million words, which make the *BSNC* larger than other specialised corpora of 19th century fiction. For instance, the *19th Century Reference Corpus* from *CLiC* (Mahlberg et al. 2016) contains c. 4.5 million words, the *HUM19UK (*Walker et al. 2019) around 13 million words, and the multi-genre corpus *CLMET3.0* (De Smet, Diller & Tyrkkö 2013) devotes c. 15 million words to the fiction sub-genre. Last, the corpus is subdivided into eight different sub-periods. This allows for diachronic analysis and, quite crucially, given the idiosyncratic nature of the corpus, for the investigation of language change across the lifespan of the eleven authors included in the corpus. In this regard, the compilers, as advocates of Diachronic Construction Grammar (Hilpert 2013; Traugott & Trousdale 2013), despite dealing with historical 'bad' (Labov 1994: 11) or 'imperfect' data (Janda & Joseph 2003: 14) are willing to (i) accept the cognitive commitment (Hilpert 2018; Petré et al. 2019; Wolk et al. 2013), (ii) adopt a usage-based approach to language (Bybee 2010), and (iii) work under the assumption that 'language users continue to fine-tune their grammars beyond childhood, and across the lifespan' (Petré et al. 2019: 84).

Several principles have guided the compilation of the corpus. First, unlike other historical corpora such as *CLMET* (De Smet 2005a, 2005b), which includes only parts of works and a maximum amount of text per author (i.e. 200,000 words), here we have opted for including full novels and going far beyond the word limit established in De Smet (2005b) for *CLMET* and the minimum amount of text per author fixed by Petré et al. (2018) for the corpus of *Early Modern Multiloquent Authors* (*EMMA*, 1623-1757, comprising 87,126,198 words, with a minimum of 500,000 words per author). In the case of male authors, we reached a total of 3,000,000 words, and in the case of female authors, 1,500,000 words (see Tables 2 and 3). In some cases, we have even included the full text of those novels that for some reason were left unfinished, such as *The Mystery of Edwin Drood* (94,960 words), which Dickens was writing when he died in 1870. These specific features of the *BSNC* allow us to trace in a more accurate way language variation and change across the lifespan of authors, as well as to establish interesting comparisons between individual-level changes and aggregate-level ones (see, for instance, Arnaud 1998; Nevalainen, Raumolin-Brunberg & Mannila 2011; Fonteyn 2017; Fonteyn & Nini 2020; Petré 2017; Petré et al. 2019; Sankoff & Blondeau 2007; Schmid & Mantlik 2015). As mentioned by

Fonteyn (2017), when investigating alternating constructions at the individual level, one way of overcoming the 'bad' data issue, especially when the absolute frequency of tokens is far from substantial, is 'to set a minimum number of tokens threshold for each author, or to make sure that a sufficiently large corpus is compiled for each author' (Fonteyn 2017: 258). By including a number of full novels for each of the eleven authors in our corpus and, even exceeding the 500,000-word maximum per author established by Petré et al. (2019), we believe that the *BSNC* will successfully overcome the problem of data that the historical linguist confronts when investigating linguistic phenomena at the aggregate and individual level, and also from a diachronic perspective (Traugott & Trousdale 2013: 41–42).

Second, as can be seen in Table 1, all the novels have been written by canonical authors. We have chosen novels from Anthony Trollope, Charles Dickens, Charlotte Brontë, Elizabeth Gaskell, George Eliot, George Meredith, Jane Austen, Mary Shelley, Thomas Hardy, Walter Scott and William Makepeace Thackeray. We are fully aware that other canonical authors could also have been included (or some of the ones selected could have been left out). However, we are positive that in this selection the authors have received some scholarly attention. In fact, as shown in Table 1, the previously mentioned 19th century corpora (i.e. *CLMET3.0* and *HUM19UK*) include sample texts from these canonical authors. This is also the case for Busse's (2010) self-compiled corpus whereby she explores discourse presentation strategies in 19th century British fiction.[1] Our corpus is therefore representative of well-read texts and makes it possible to establish connections with, for instance, the findings obtained from other previous studies based on these other diachronic corpora.

A third criterion has to do with the period of publication of the novels and the author's year of birth. As represented in Figure 1, we distinguished three main literary periods, and the texts included in each of them (i.e. Romanticism, Victorian, and Aestheticism and Decadence) were written by authors born within a correspondingly restricted time-

[1] Mahlberg et al.'s (2016) *19th Century Reference Corpus* is not included in Table 1, as it was not compiled to be representative of 19th century English fiction in general, but as a comparable collection of texts to Dickens's works (see Mahlberg 2013: 42–43).

span. Thus, we have Generation 1 (1771-1799), Generation 2 (1800-1820), and Generation 3 (1821-1850). As in De Smet (2005b), the purpose was 'to increase the homogeneity' within the parts whereby '[h]istorical trends […] appear somewhat more clearly' (2005b: 70). Except for Trollope (1815-1882), from Generation 2, and Meredith (1828–1909) and Hardy (1840-1928), from Generation 3 (1821-1850), no author's work is represented in two subsequent literary periods.

Table 1. Canonical authors represented in the *BSNC* and other 19th century fiction corpora

Canonical authors	***BSNC***	***CLMET3.0***	***HUM19UK***	**Busse (2010)**
Austen	✓	✓	✓	✓
Scott	✓	✓	✓	✓
Shelley	✓	✓	✓	✓
Dickens	✓	✓	✓	✓
Thackeray	✓	✓	✓	✓
Brontë	✓	✓	✓	✓
Gaskell	✓	✓	✓	✓
Trollope	✓	×	✓	×
Meredith	✓	✓	✓	×
Eliot	✓	×	✓	✓
Hardy	✓	✓	✓	✓

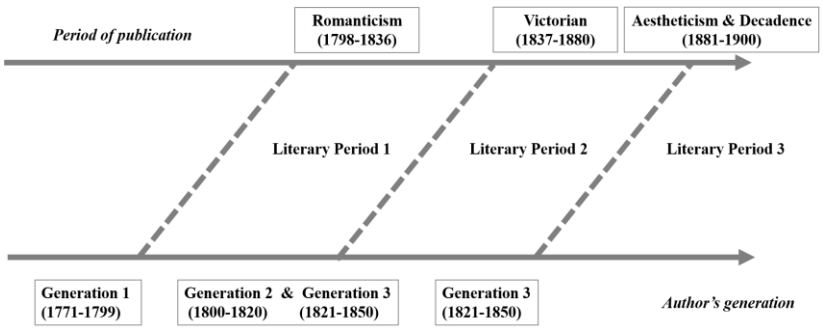

Figure 1: Correspondence between the novels' period of publication and the author's generation

Fourth, all authors are native speakers of British English. This measure aimed at controlling for dialectal variation and facilitate comparisons with other historical and Present Day English corpora based on both British and American English (e.g. *ARCHER-3.2*, *CLMET3.0*, *CEN*). As shown in one of our own case studies, this criterion proved to be crucial to be able to confirm previous findings on the history of the ROC, a valency-increasing construction hypothesised to be a British innovation that later spread over to the American variety (Bouso 2021).

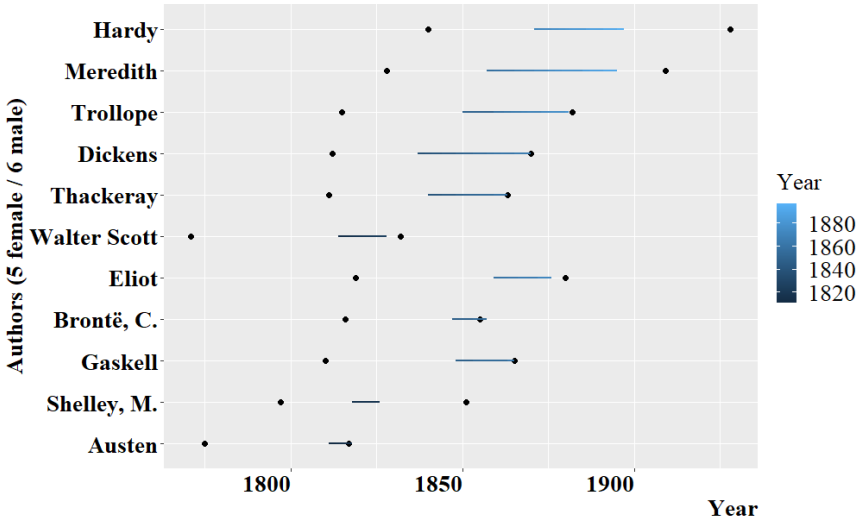

Figure 2. Lifespan and active career of the *BSNC* authors as represented in the corpus; the lines represent the authors' active career and the two dots the author's date of birth and death, respectively.

Fifth, in our selection of authors we have tried to keep a balance between male and female writers. As can be seen in Figure 2 (from bottom up, from older to newer generations), there are five female novelists (Jane Austen, Mary Shelley, Elizabeth Gaskell, Charlotte Brontë, and George Eliot) and six male novelists (Walter Scott, William Thackeray, Charles Dickens, Anthony Trollope, George Meredith, and Thomas Hardy). Two of the female writers belong to the Romantic period (Austen and Shelley) whereas the other three (Gaskell, Brontë and Gaskell) belong to the Victorian era. No female writers have been included yet for the Aestheticism and Decadence period (1881–1900).

We had wished to consider here some variation regarding the social background of the authors along the lines of *EMMA* (Petré et al. 2019). However, at this stage this has not been possible since the eleven canonical authors selected clearly belong to the higher echelons of 18th and 19th century English society, as this happens to be the case in the vast majority of the texts included in *Project Gutenberg* (De Smet 2005b: 71).

The gender parameter has determined the sixth criterion, i.e. the number of novels included per author. In this regard, it is important to bear in mind that female authors were less prolific than men, partly because they were 'influenced/shaped/restricted' by the society they lived in (Šalinović 2014: 218). So much so that an author like Mary Anne Evan decided to publish under the male pen name of George Eliot to get more credit (Bodenheimer 1994). This socio-historical context explains why the number of novels per author is not the same for both genders nor even for the three main literary periods in which the corpus is subdivided. As shown in Tables 2 and 3 respectively, for women the number of novels ranges from 3 to 7, and for men from 12 to a maximum of 19. It should be noted that some male authors wrote more than 19 novels (Walter Scott and Anthony Trollope, for instance). However, we decided to include fewer than 20 novels per author in order to keep a more balanced distribution of texts than if every novel by every author had been included in the corpus.

Table 2: Distribution of the novels in the *BSNC* across female authors

Authors	Novels	Words
Austen	6	743,288
Brontë	4	691,159
Eliot	7	1,544,808
Gaskell	6	1,048,185
Shelley	3	309,086
	26	**4,336,526**

Table 3: Distribution of the novels in the *BSNC* across male authors

Authors	Novels	Words
Dickens	15	3,874,658
Hardy	14	1,726,956
Meredith	14	2,033,362
Thackeray	12	2,086,993
Trollope	14	3,527,797
Scott	19	3,565,828
	88	**16,815,594**

Table 4. Internal structure of the *BSNC* (an asterisk indicates authors represented in two successive periods)

	Authors	Novels (per sub-period)	Words (per sub-period)
Generation 1 (1771-1799) **Romantic Period** (1798-1836)	***BSNC1* (1798-1820)**	**17**	**2,584,897**
	Austen	6	743,288
	Shelley	2	129,552
	Scott	9	1,712,057
	***BSNC2* (1821-1836)**	**11**	**2,033,305**
	Shelley	1	179,534
	Scott	10	1,853,771
Total number of words Generation 1: 4,618,202			
Generation 2 (1800-1820) **Victorian Period** (1837-1880)	***BSNC3* (1837-1850) - Early Victorian**	**19**	**4,051,335**
	Brontë	2	405,247
	Dickens	8	2,323,774
	Gaskell	1	162,297
	Thackeray	7	980,511
	Trollope	1	179,506
	***BSNC4* (1851-1960) - High Victorian**	**18**	**3,284,005**
	Brontë	2	285,912
	Dickens	4	940,747
	Eliot	2	426,392

	Gaskell	3	418,972
	*Meredith	2	205,696
	Thackeray	4	983,759
	Trollope	1	22,527
	***BSNC5* (1861-1870) - High Victorian**	**18**	**4,059,461**
	Dickens	3	610,137
	Eliot	3	485,050
	Gaskell	2	466,916
	*Meredith	3	526,992
	Thackeray	1	122,723
	*Trollope	6	1,847,643
	***BSNC6* (1871-1880) - Late Victorian**	**19**	**3,545,750**
	Eliot	2	633,366
	*Hardy	7	875,090
	*Meredith	5	637,449
	*Trollope	5	1,399,845
Total number of words Generation 2: 14,940,551			
Generation 3 (1821-1850) **Aestheticism and Decadence Period** (1881-1900)	***BSNC7* (1881-1890)**	**7**	**911,120**
	*Hardy	4	490,182
	*Meredith	2	342,662
	*Trollope	1	78,276
	***BSNC8* (1891-1900)**	**5**	**682,247**
	*Hardy	3	361,684
	*Meredith	2	320,563
Total number of words Generation 3: 1,593,367			

Finally, the number of novels per author was also dependent on our last and perhaps most important criterion, i.e. their year of publication. As shown in Table 4, we have distinguished eight sub-periods in total, distributed across the three generations established. We tried to include in each sub-period novels of female and male authors, and to achieve an approximate equal distribution in the number of texts and words. This division of the corpus into eight sub-periods of roughly 10 years each makes it possible to establish comparisons with the other historical

corpora previously mentioned and also to fine-tune our diachronic linguistic analyses, aiming at keeping track of 'gradual' and/or 'small-step' abrupt language changes (Traugott & Trousdale 2013), and investigating stylistic and morpho-syntactic variation across the lifespan of authors, especially those authors whose complete novelistic trajectory has been covered in full, as in the case of Elizabeth Gaskell, George Eliot or Charles Dickens.

3. Research Possibilities and Examples

3.1 Previous Studies

In this section we review some of the research in which the *BSNC* has been used. It must be clear that in most studies conducted the data retrieved from the *BSNC* has been supplemented with data from other historical corpora. As will be shown below, initially the corpus was mostly used in studies involving qualitative analysis, and more recently it has also been proven successful in quantitative studies concerning language change. Still, we are fully aware of the fact that the corpus needs to be further tested to identify strengths and weaknesses and ways to overcome the latter.

Previous studies such as those by Ruano San Segundo (2016a, 2016b, 2018a, 2018b) have used a preliminary version of the *BSNC* to identify reporting verbs in 19th century fictional narratives and to test the hypothesis that these verbs are used as a characterising device in Charles Dickens's novels. Ruano San Segundo (2018b) shows that verbs that describe sounds made by animals (*growl*, *roar*, etc.) are frequently used by Dickens to report the words of villains. Differences based on character gender have also been identified. Thus, verbs like *thunder* or *vociferate* are exclusively associated with men in Dickens's narrative, whereas choices like *pout* or *moan* tend to be used to gloss female discourse (Ruano San Segundo 2018a). These findings have contributed to further reinforcing Dickens's well known hyperbolic style (Gomme 1978: 72) and to expanding the analysis of his celebrated techniques of characterisation as far as the individualisation of speeches is concerned (Golding 1985), one of the aspects for which he is best known.

More recently, Bouso (2020b) used the visualisation tool of 'animated' motion charts (Hilpert 2011; Hilpert & Perek 2015) to test the 19th century diversity of the ROC, where an originally intransitive verb of manner of action (e.g. *smile*) is followed by a non-prototypical type of

object that describes a reaction or an emotion of some kind (e.g. *disbelief*), as in the prototypical ROC example *She smiled disbelief*. Building on these findings, and other previous research on the topic conducted by one of the authors (Bouso 2017, 2020a, 2021), Bouso and Ruano San Segundo (2021) explored, on the one hand, the tight relation between the ROC and the 19th century British sentimental novel and, on the other, the role of Direct Discourse Constructions (DDCs) of the type *She smiled, 'I don't believe you'* in the development of the ROC.

On the whole, the results from these studies (i.e. Bouso 2020b; Bouso and Ruano San Segundo 2021) concur with Bouso's (2021) hypothesis that the proliferation of the ROC could well be a direct consequence of the continuous development of the novel and, most particularly, of the British sentimental novel. To be more precise, Bouso and Ruano San Segundo's (2021) results indicate that the ROC is especially frequent in the *BSNC* by comparison, for instance, with analogous data retrieved from the American section of the *CEN* (1881-1922) or the multi-genre corpus *CLMET3.0* (1710-1920). Also, the highest peak in type frequency of the ROC falls right in the middle of the 19th century (the *BSNC3* sub-period corresponding to 1837-1950), which coincides with what is considered to be the heyday period of the British sentimental novel (Williams 2020: 20). As for the second research question discussed in Bouso and Ruano San Segundo (2021)—that concerning the role of the DDC in the development of the ROC—, the *BSNC* and *CEN* data reveals a significant strong positive time-frequency correlation in the development of the ROC and the DDC.

Importantly, Bouso and Ruano San Segundo (2021) also introduce in their paper the novel idea that the ROC and the DDC are functionally similar constructions since they occur with manner of action verbs such as *smile* and *nod* in (1) and (2), and both 'can be treated as one more option to project someone else's verbal or mental discourse' (2021).

(1) a. 'Well,' said Mr. Pickwick; 'but I must send a letter to London by some conveyance, so that it may be delivered the very first thing in the morning, or I must go forwards at all hazards.' The landlord *smiled* ***his delight***. *Nothing could be easier* than for the gentleman to inclose a letter in a sheet of brown paper, and send it on, either by the mail or the night coach from Birmingham. (*BSNC* 1836-37, Dickens; *The Pickwick Papers*)

b. He is a very creditable artist himself. *He will be delighted,* I am sure, with Mrs Granger's taste and skill.' 'Damme, Sir!' cried Major Bagstock, 'my opinion is, that you're the admirable Carker, and can do anything.' ***'Oh!'*** *smiled* Carker, with humility, *'you are much too sanguine, Major Bagstock.'* (*BSNC* 1846-48, Dickens; *Dombey and Son*)

(2) a. It sympathized on the side of his backers too much to do more than *nod* ***a short approval*** of his fortitude. (*BSNC* 1895, Meredith; *The Amazing Marriage)*

b. 'The prince is a gentleman, grandada. Come with me. We will go alone. You can relieve the prince, and protect him.' My father *nodded*: ***'I approve.'*** (*BSNC* 1870-71, Meredith; *The Adventures of Harry Richmond*)

To justify the functional similarity between the ROC (in e.g. (1a) and (2a)) and the DDC (in e.g. (1b) and (2b)), Bouso and Ruano San Segundo (2021) relied on the model developed by Semino & Short (2004) which distinguishes the following categories of discourse presentation constructions: Narrative Report of Speech Acts (e.g. *He told her about* ***his imminent return***), Indirect Speech (e.g. *He told her* ***that he would definitely return the following day***), Free Indirect Speech (***He would definitely come back tomorrow!***), Direct Speech (*He said* ***'I'll definitely come back tomorrow!'***) and Free Direct Speech (***'I'll definitely come back tomorrow!'***). These categories differ from each other in that they 'are associated with different degrees of faithfulness to an original' (2021). In other words, by moving along the cline, from the more bound (i.e. Narrative Report of Speech Acts) to the more free end (i.e. Free Direct Speech), the interference of the reporter's part becomes less and less noticeable. In this cline, the ROC, as represented in (1a), fits into the more bound category of NRSA (Narrative Report of Speech Acts) (e.g. *He told her about* ***his imminent return***) 'where a minimal account of the discourse reported is given' (Bouso and Ruano San Segundo 2021). In this category a reporting verb (i.e. *told*) is followed by a Noun Phrase that indicates the topic of the speech (or thought) presented (i.e. *imminent return*). In the ROC, more than just merely

indicating the topic, there is also a clear motivation on the part of the reporter to account for the illocutionary force of the speaker's utterance (e.g. joy, approbation, discontent, etc.). Though formally similar, one could say that in the ROC 'the narrator is closer [than in the NRSA] to the original act of communication conveyed by the character' (Bouso and Ruano San Segundo 2021). The motivation of the ROC to represent the illocutionary force of the speaker's original utterance becomes even more clear when comparing ROC-DDC alternations in (1) and (2). In both cases the illocutionary force of the message is conveyed (i.e. the character's feeling of delight in this case); the difference is that while the DDC is more free from the narrator and therefore more character-oriented and 'faithful to an original', the ROC is clearly more subjective as it relies on the narrator's interpretation of the illocutionary force of the communicative scene it evokes.

The qualitative analysis of the aggregate data from the *BSNC* allowed the authors of the paper not only to identify such functional similarity between the ROC and the DDC but also to develop further the idea that the ROC could have been used by 19th century authors as an 'extravagant' (Haspelmath 1999), more noticeable and subjective alternative to the DDC. Finally, it should be noted that this study where the ROC is presented as a 19th century 'extravagant' alternative to the DDC ties in well with Bouso's (2021) earlier research that places the consolidation, or 'grammatical constructionalisation' (Traugott & Trousdale 2013: 124–125) of the ROC precisely in the transition from the 18th to the 19th century. Here one could draw an interesting comparison with Petré's (2017) pioneering study which, on the basis of a number of contextual cues, provides empirical evidence of the role of 'extravagance' as an important cognitive motivation in the early stages of the grammatical constructionalisation of the progressive (e.g. *'Dennis what are you doing?' 'I'm eating because I'm very hungry.'*) (Petré 2017: 230). In the following section, we present a new case study that builds on these previous quantitative and qualitative findings on the use of the British ROC.

3.2 A Case Study

In this section, we showcase the suitability of the *BSNC* for exploring some of the motivations underlying language variation and change at the level of individuals. As mentioned by Petré et al. (2019: 251), the

practice of studying language on the aggregate level 'has dominated the methodologies in historical linguistic studies, and very little attention is paid to the individual level' when it is precisely 'the repeated behaviour of individual users' that drives linguistic change (Petré 2017; Fonteyn 2017; see also Traugott & Trousdale 2013; Trousdale 2019: 12). In addition to this, Fonteyn (2017: 255) observes that 'it is not necessarily the case that aggregate-level tendencies accurately (or even roughly) reflect the behaviour of individual language users'. The present study contributes to addressing this lack of attention given to the individual and tests whether our previous aggregate findings still hold at the individual level. To be more precise, aggregate data from the *BSNC* indicates that the ROC (i.e. *The landlord smiled* ***his delight*** in (1a)) and the DDC (i.e. ***'Oh!'*** *smiled Carker*, in (1b)) have been competing over the same functional environments since around the 19th century (1798-1900). This is the period in which both constructions develop in a significant parallel, correlated fashion, with the ROC presumably being used as an 'extravagant' alternative to the less 'expressive' and older variant DDC (Cichosz 2018). In this light, specific research questions that will be the focus of attention here are:

RQ1: **a.** Firstly, do these competing constructions also correlate at the level of individuals?

b. And if so, are there any significant changes in this correlation across their lifespan? To phrase it differently, do these individuals, (i.e. the eleven canonical authors that comprise the corpus) become more 'conservative' or more 'progressive' over the years (Nevalainen, Raumolin-Brunberg & Mannila 2011)?

RQ2: And, secondly, does the ROC-DDC alternation represent an actual cognitive reality for individual language users, namely for these individual writers?

As aptly indicated by Fonteyn (2017), in the literature on the role of competition in linguistic change (Traugott & Trousdale 2013: 18), the functional-semantic overlap between two forms can have two possible outcomes: (i) retention and niche formation, or (ii) replacement (or substitution, with one of them declining). In cases of retention, the

competing constructions develop towards a division of labour, with each of the constructions preferred in particular functional niches (see Torres Cacoullos & Walker 2009). To illustrate how this division of labour may work at the level of individual language users, Fonteyn (2017: 255) proposes three hypothetical scenarios (see Figure 3). In Scenario 1, 'all individuals use the alternation' (Fonteyn 2017: 255). In Scenario 2, the 'division of labour only represents the behaviour of a part of the entire population' (Fonteyn 2017: 255), that is, one of the variants (the new or the older one) is used by a small group of individuals. In the last hypothetical scenario, all individuals can be either 'progressive' or 'conservative' (Nevalainen, Raumolin-Brunberg & Mannila 2011), depending on whether they use the new or the old variant, respectively. In this last case scenario, following Petré (2017), Fonteyn (2017: 255) notes that 'the functional-semantic division of labour observed on the aggregate level only exists in the language as an abstract object'.

In what follows, we will attempt to segregate part of the aggregate data included in Bouso and Ruano San Segundo (2021). The data used here has been drawn exclusively from the *BSNC*, which, as shown in Table 4, is divided into eight sub-periods. For each sub-period we retrieved all tokens of DDCs and ROCs for the seven most prototypical verbs of the Late Modern English (LModE) ROC, previously identified via a distinctive collexeme analysis (Stefanowitsch and Gries 2003) by one of the authors of the paper (Bouso 2017). These are the verbs *mutter*, *murmur*, *smile*, *nod*, *whisper*, *shout* and *wave*. The data obtained was then manually pruned, filtering out those instances that did not count as prototypical examples of the ROC or the DDC (for details, see Bouso and Ruano San Segundo 2021). A total of 434 ROCs and 2920 DDCs were retrieved from the 114 novels that form the corpus. The focus of attention was on the ROC-DDC alternation as represented through the *BSNC* authors' novels as these can be seen as a window into the cognitive reality of the alternation in the authors' minds at the time of their publication (Schmid & Mantlik 2015: 585). Finally, before moving into the results of our study, it should be mentioned that all statistical analyses are based on (standardised) normalised frequencies per million words and correlations were established on the basis of Pearson's product-moment coefficient *r* (Levshina 2015).

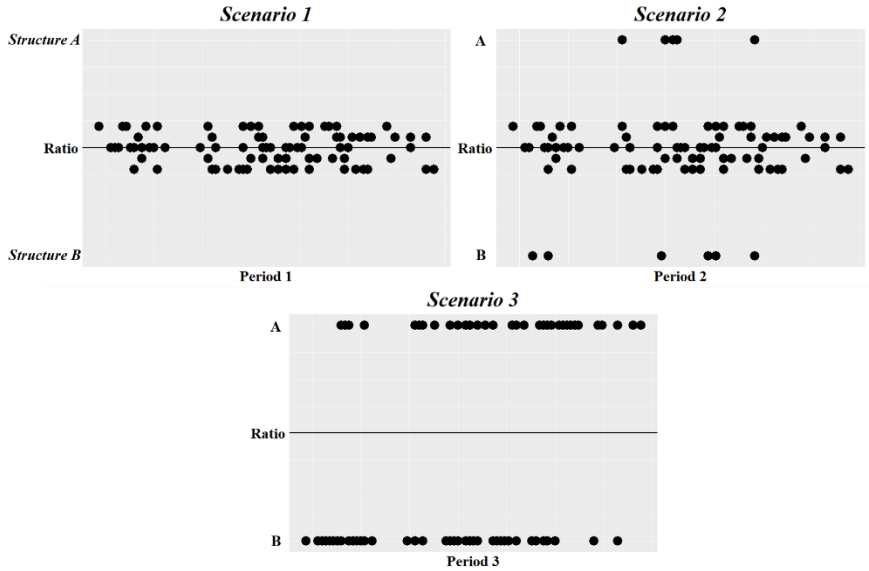

Figure 3: Hypothetical scenarios of alternation between two competing structures. The dots represent the alternating behaviour of individuals (adapted from Fonteyn 2017: 255).

As noted in section 3.1, aggregate-level data from Bouso and Ruano San Segundo (2021) evince a significant strong positive time-frequency correlation between the ROC and the DDC over the course of the 19th century and the first two decades of the 20th century. Temporal correlations of this type, however, must be examined carefully in order to be affirmed as 'systematic' (Flach 2013; Hilpert 2020; Liberman 2013). Koplenig & Müller-Spitzer (2016), for instance, claim that 'whenever two variables evolve through time, those variables will almost always look highly correlated even if they are not related in any substantial sense' (2016: 2). To illustrate this idea, they provide as examples the correlations that exist between population size and lexical diversity on the one hand, and the mean sea levels and lexical diversity on the other. The strong and highly significant correlations they instantiate would mistakenly lead the analyst to think that both (i.e. population size and mean sea level) can, in fact, predict lexical diversity when it is clear that the latter is a 'nonsensical' or spurious type of correlation. Following Koplenig & Müller-Spitzer (2016), Hilpert (2020) proposes a possible solution: we should accept correlations only if individual changes

correlate; these individual changes would 'act as a powerful filter against potentially spurious correlations' (Hilpert 2020: 12).

The aggregate-level data from Bouso and Ruano San Segundo (2021) when examined from this perspective reveals that the DDC and the ROC are clearly similar with regard to their individual changes. As shown in Figure 4, but for the period between 1871 and 1880 (*BSNC6*), individual changes correlate across the entire trend in such a way that when the ROC goes up the DDC also goes up and vice versa. Note here that points above 0 indicate an increase in frequency between two time frames, and points below 0 indicate a decrease in frequency. The last graph in Figure 4 (i.e. ROC / DDC) further confirms our previous results as it shows that any change in the DDC (independent variable) for a given period strongly predicts the kind of change that will be observed in the ROC (dependent variable) during that time ($p < 0.05$). The changes are not completely identical, but clearly they go side by side 'in terms of direction and magnitude' (Hilpert 2020: 12).

If we now pay attention to the ROC-DDC correlation at the level of the individual (RQ1-a), a mixed, hybrid picture emerges. Figure 5 plots ROC-DDC correlations for the eleven authors in the *BSNC*. All authors alternate between the ROC and the DDC in the vast majority of their novels. However, the correlation only turns out to be statistically significant for Walter Scott ($p < 0.05$) and Charles Dickens ($p < 0.05$)[2] (RQ1-a) and no significant changes in their use of the ROC-DDC alternation have been detected across their lifespans ($p > 0.05$) (data not shown). This indicates that, as far as this alternation is concerned, both Scott and Dickens remain faithful to their own narrative style, not becoming either more 'progressive' or 'conservative' (Nevalainen, Raumolin-Brunberg & Mannila 2011) over the years (RQ1-b).

[2] A larger set of novels by Austen and Shelley would have been desirable to be able to reach conclusive results for these authors as well (NA = correlation coefficient was not available because of insufficient data).

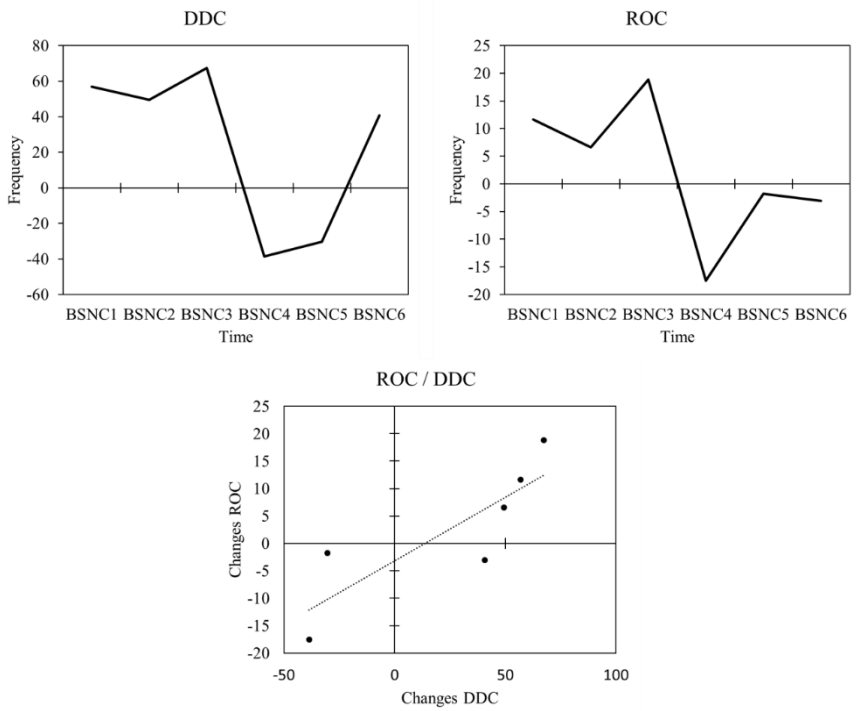

Figure 4: Frequency developments in the DDC and the ROC. In the third panel, the *x*-axis represents the DDC and the *y*-axis the ROC ($p < 0.05$). The last two sub-periods of the *BSNC* (1881-1900) have been excluded to avoid skewing the results; as shown in Table 4 (section 2), these periods are somewhat underrepresented with regard to the other sub-periods of the corpus.

Moving on to the last research question concerning the cognitive reality of the ROC-DDC alternation in the author's minds (RQ2), Table 5 lists in the columns shaded in grey the number and percentage of novels in each sub-period that used both ROCs and DDCs (alternating novels) and the number and percentage of novels in each sub-period that only used one of the two types (non-alternating novels). Figure 6 zooms in on Table 5 and plots the alternating behaviour (per individual novel) over the course of the 19th century. The Romantic period (*BSNC1-BSNC2*; 1798-1836) is marked by a balanced use of DDCs and ROCs. On the other hand, the Victorian period (*BSNC3-BSNC6*; 1837-1880) makes extensive use of the default discourse presentation strategy (i.e. the DDC), with the ROC standing its ground in a small usage niche. Finally,

the last period, Aestheticism and Decadence (1881-1900), is characterised by a ROC retention despite considerable losses. To sum up, in the three hypothetical retention scenarios (Figure 3) proposed by Fonteyn (2017), the *BSNC* individual data fits best within Scenario 2: the division of labour between the DDC and the ROC in the aggregate data represents the vast majority of the *BSNC* authors' novels and only a relatively small group (17.5%) appears to use one of the variants. This is more explicitly shown in Figure 7.

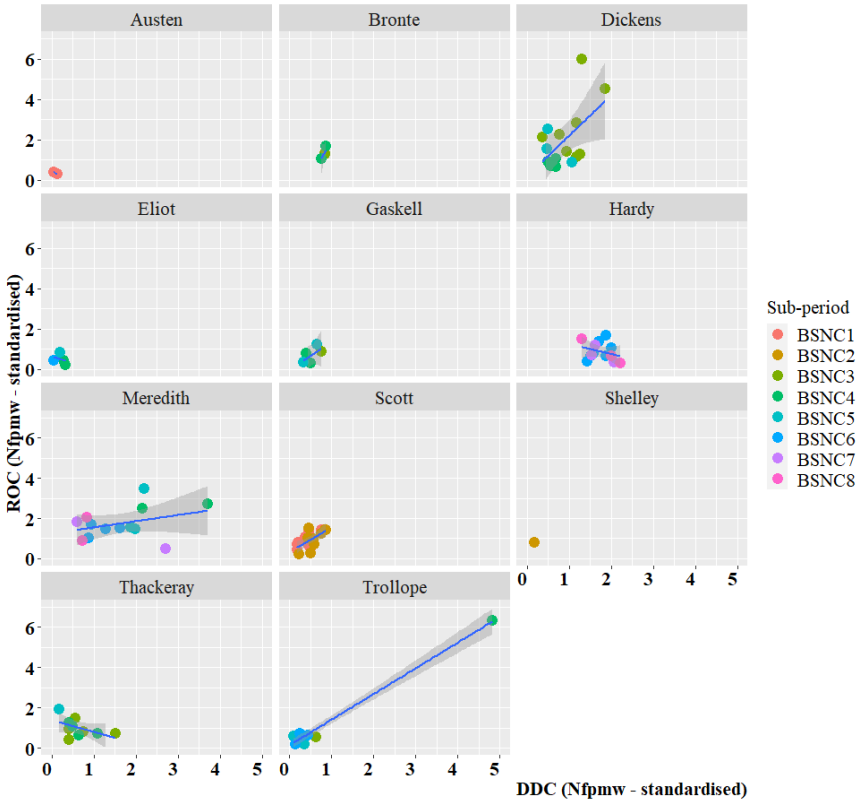

Figure 5: ROC-DDC correlations in (standardised) normalised frequencies per million words for the 11 alternating authors of the *BSNC* (in alphabetical order). The correlation is statistically significant for Dickens ($p < 0.05$) and Scott ($p < 0.05$). For Trollope the correlation was significant ($p < 0.001$), but it turned out to be non-significant when the outlier (green point) was removed ($p > 0.05$). Non-alternating novels have been excluded. Dots represent the author's novels and the colour the corresponding period in which each novel was published.

Table 5. Frequencies of novels per sub-period

	Number of occurrences of the ROC and the DDC per sub-period		**Novels total**	**Alter-nating (ROC-DDC)**	Force of ROC-DDC freq-uency corre-lation per sub-period[3] (alter-nating)	**Non-alter-nating total**	Only ROC	Only DDC
	ROC	**DDC**						
BSNC 1	30	147	17 (of 17)	10 (58.8 %)	0.52*	7 (41.1%)	0	7
BSNC 2	37	216	11 (of 11)	10 (90.9%)	0.30	1 (9.09%)	0	1
BSNC 3	150	703	19 (of 19)	17 (89.4%)	0.23*	2 (10.5%)	0	2
BSNC 4	64	443	18 (of 18)	17 (94.4%)	**0.85*** **	1 (5.5%)	0	1
BSNC 5	72	424	18 (of 18)	14 (77.7%)	0.27*	4 (22.2%)	1	3
BSNC 6	52	515	19 (of 19)	16 (84.2%)	0.22*	3 (15.7%)	0	3
BSNC 7	14	286	7 (of 7)	5 (71.4%)	0.65	2 (28.5%)	0	2
BSNC 8	15	186	5 (of 5)	5 (100%)	0.37	0 (0%)	0	0
Total	**434**	**2920**	**114**	**94 (82.4%)**		**20 (17.5)**	**1**	**19**

N.B. Force of correlation (Pearson's *r* with the associated *p*-values): *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

[3] The ROC-DDC frequency correlation is highly significant ($p < 0.001$) for the 19th century data (including alternating novels; excluding non-alternating ones).

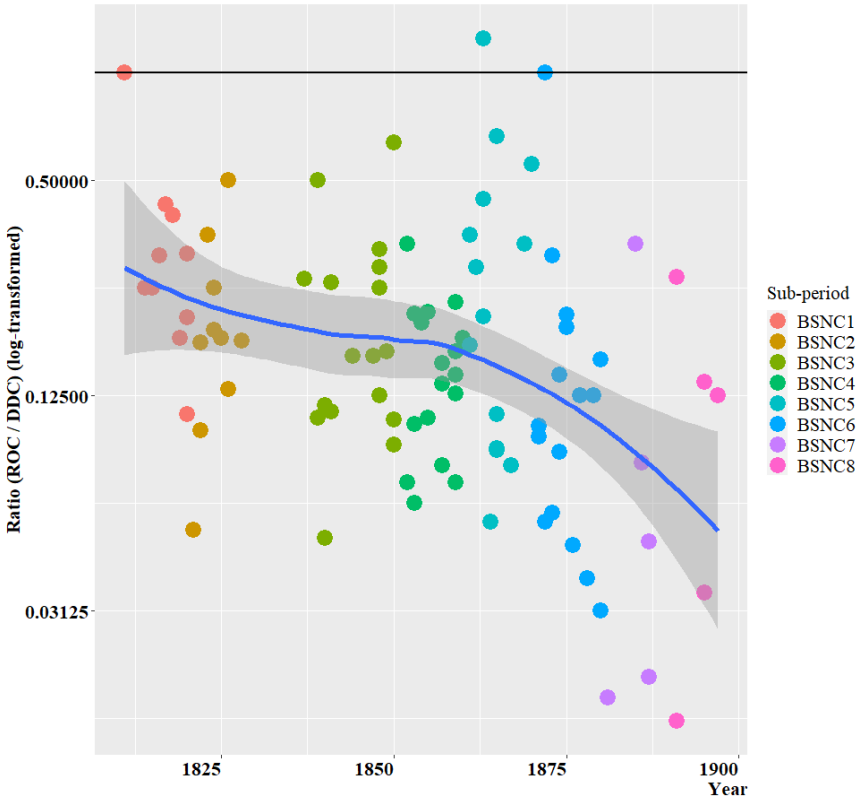

Figure 6: Diachronic development of the ROC-DDC alternation. The LOESS smoothed curve shows the general direction of the alternation (Levshina 2015) at the aggregate level. Dots towards the bottom of the graph represent novels that are more inclined to use the DDC, whereas dots towards the top of the graph represent novels that lean towards the use of the ROC. Dots that fall on the black line show an equal number of both constructions. Non-alternating novels have been excluded.

Figure 7 shows how the non-alternating novels that use the more conservative variant (i.e. the DDC) are sparingly distributed throughout the three literary periods of the 19th century. In the first period, conservative novels—or novels that only use the older and less expressive DDC—account for roughly 30% of the data. This percentage considerably decreases during the Victorian era (1837-1880) when the ROC undergoes a process of niche formation and stabilisation. As shown in Figure 6, over this literary period, (i) the number of dots near the black

line is noticeable, and (ii) the force of the ROC-DDC correlation reaches its highest point, namely between 1851 and 1860 (see Table 5). This Golden Age of the ROC ends with the Aestheticism and Decadence period (1980-1900) when a 'process of decoupling' (De la Pena & Giné 2012) is produced and the more conservative DDC starts to gain ground again at the expenses of the ROC (see force of correlation for these two sub-periods in Table 5). This process of decoupling—or lack of statistical correlation between these two variables—may be due to a number of factors, among which are the gradual process of independence of the ROC from the DDC and its resulting loss of extravagant power. Since the ROC is no longer used merely as an alternative to the DDC, its extravagant dimension may no longer be perceived as such by the readers of these novels.

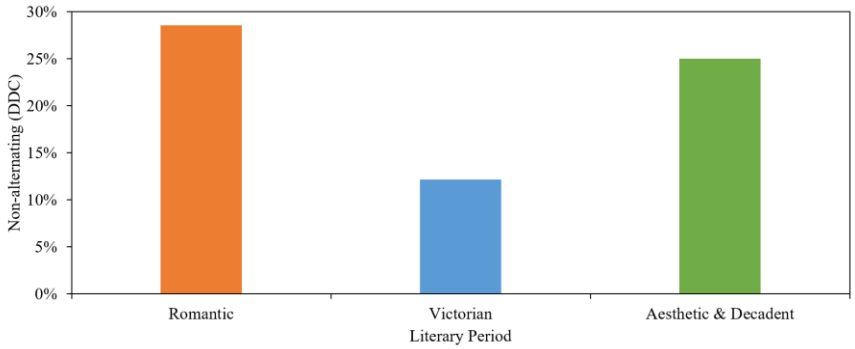

Figure 7: Diachronic changes in the percentage of non-alternating novels using solely the DDC. N.B. Table 5 provides the percentage of non-alternating novels per sub-period.

To elaborate on this, Bouso (2020a, 2021) shows, on the basis of historical data retrieved from the *Corpus of Historical American English* (*COHA*, see Davies 2010), that at the turn of the 20th century the ROC starts to expand via different mechanisms of change to a number of different verb types and classes. At the start of the 20th century, the ROC is mostly dominated by the verbs of manner of action analysed here, and as we move into the 20th century, verbs of instrument of communication (*phone* as in *He* ***phoned*** *good-bye to some neighbours*), verbs of activity (*play* as in *Tonight we* ***played*** *goodbye to the piano*), verbs of light emission (*flare* as in *The star* ***flared*** *its good-bye*), and even very low-frequency verbs of manner of action (*smooch* as in *Our morning routine*

*of **smooching** good-bye at the train station had somehow fallen by the wayside*) enter the ROC via coercion, metaphor and analogization with the verbs functioning as prototypes during this period, for instance, the verb *kiss*, which is attested in semi-fixed collocations such as *kiss goodbye* (cf. Bouso 2021, Martínez-Vázquez 2014). As shown in the ungrammatical examples in (3), these new instantiations of the 20th century ROC are hard to see as part of the ROC-DDC alternation in the way exemplified in (1) and (2) (section 3.1). Quite interestingly, this gradual loss of extravagant power of the ROC during the Aesthetic and Decadent period seems to be the natural path of development (Haspelmath 1999), as also recently shown, for instance, in Petré's (2017) account of the history of the progressive, where he also distinguishes two stages in its development: one first extravagant stage which is distinctive of a first generation of writers, and a second stage that is characterised by a conventionalisation process and a 'decrease of extravagance' (Petré 2017: 247).

(3) a. *He **phoned**, 'Good-bye'.
b. *We **played**, 'Good-bye'.
c. *The star **flared**, 'Good-bye'.
d. *We **smooched**, 'Good-bye'.

3.3 ROC and Style

In this last section, we briefly discuss stylistic implications of the ROC from a literary point of view. To do so, we focus on Charles Dickens, who stands out from the rest of the authors in the *BSNC* for his use of both DDCs and ROCs. More specifically, he accounts for more than 20% of the total of DDCs (682/2920) and more than 35% of the overall ROCs (152/434). DDCs have traditionally been regarded as one of Dickens's best-known techniques of characterisation to depict characters' memorable voices. First, direct speech reported clauses are used to render individualized voices through the use of catchphrases, dialectal varieties and other speech peculiarities. Second, elements in direct speech reporting clauses, such as reporting verbs, serve to reinforce the depiction of these voices. As mentioned in section 3.1, certain direct speech descriptive verbs (Caldas-Coulthard 1987: 162) are used exclusively with characters of the same gender, while others are associated with similar character types. This is the case for male villains,

whose speech is frequently introduced by verbs that describe sounds made by animals (*bawl*, *growl*, *croak*, etc.) (see Ruano San Segundo 2018b). Interestingly, less conspicuous reporting verbs are also meaningful in terms of characterisation. Let us take Ralph Nickleby (*Nicholas Nickleby*), Bill Sikes (*Oliver Twist*) and Daniel Quilp (*The Old Curiosity Shop*) as examples. Not only is the speech of these three well-known Dickensian villains shaped by verbs such as *bawl*, *growl* and *croak*, but their evil nature is also projected with less noticeable choices, such as *mutter*. Examples (4a–c) serve to illustrate this idea. These examples could seem coincidental, but they are not. In the case of Ralph, for instance, *mutter* is used sixteen times to report his words in *Nicholas Nickleby* (Ruano San Segundo 2017: 115). Besides this, the reporting verbs are frequently accompanied by an adjunct that somehow reinforces the characterizing role of the reporting verb (Ralph mutters sternly, Sikes mutters grinding his teeth and Quilp mutters darting an angry look at his wife).

(4) a. 'I am not a man to be moved by a pretty face,' ***muttered*** Ralph sternly. 'There is a grinning skull beneath it, and men like me who look and work below the surface see that, and not its delicate covering.' (*BSNC* 1837-39, Dickens; *Nicholas Nickleby*)

b. 'Wolves tear your throats!' ***muttered*** Sikes, grinding his teeth. 'I wish I was among some of you; you'd howl the hoarser for it.' (*BSNC* 1838-39, Dickens; *Oliver Twist*)

c. 'Humph!' ***muttered*** the dwarf, darting an angry look at his wife, 'I thought it was your fault! And you, sir—don't you know there has been somebody ill here, that you knock as if you'd beat the door down?' (*BSNC* 1840-41, Dickens; *The Old Curiosity Shop*)

This widely discussed characterising role of reporting verbs in DDC can also be extended to ROCs. There exist, in fact, striking similarities in the use of the same verbs in the DDC and the ROC. Thus, *mutter* is also used to report the words of Ralph, Sikes and Quilp in ROCs, as shown in (5a–c), respectively. Moreover, they all have the same type of reaction object: a deverbal illocutionary noun (mutter a peevish interjection,

mutter a curse and mutter a terrible oath), which contributes to shaping the evilness of the characters in the same manner that glossing phrases in examples in (4a–c) do in DDCs. These examples suggest that Dickens maintains the characterising patterns with which characters are depicted when he opts for ROCs. This is an aspect hitherto underexplored in literary appreciations of his style, and obviously requires a more comprehensive analysis than editorial constraints permit here. However, in light of these examples (same verb with same character in both DDCs and ROCs), the literary use of the ROC can hardly be dismissed as a coincidence, but it seems another technical device among Dickens's well-known techniques of characterisation (cf. Paroissien 2000). This literary use is actually worthy of further analysis, as it could lead to a further understanding of the widespread use of the ROC in the British sentimental novel in general and in the craftsmanship of specific authors in particular.

(5) a. At such intervals, after a few moments of abstraction, Ralph would *mutter* ***some peevish interjection***, and apply himself with renewed steadiness of purpose to the ledger before him, but again and again the same train of thought came back despite all his efforts to prevent it, confusing him in his calculations, and utterly distracting his attention from the figures over which he bent. (*BSNC* 1837-39, Dickens; *Nicholas Nickleby*)

b. Illness had not improved Mr. Sikes's temper; for, as the girl raised him up and led him to a chair, he *muttered* ***various curses*** on her awkwardness, and struck her. (*BSNC* 1838-39, Dickens; *Oliver Twist*)

c. The dwarf *muttering* ***a terrible oath*** looked round as if for some weapon with which to inflict condign punishment upon his disobedient wife. (*BSNC* 1840-41, Dickens; *The Old Curiosity Shop*)

To conclude, this case study has aimed to test the validity and reliability of the *BSNC* to conduct qualitative and quantitative studies at the individual level. It adjusts to the two items in the research agenda of the cognitive historical linguists mentioned by Fonteyn (2017) and that

have recently become the focus of attention in the field (see also Fonteyn & Nini 2020: 280; and other recent articles directly stemming from Petré's (2015-2021) ERC-funded project *Mind-Bending Grammars*). These two items concern the comparison between aggregate and non-aggregate data and the exploration of non-alternating individuals where only one of the variants is attested. This was done in such a way as to maximally exploit the historical data and achieve an in-depth view of the historical development of the LModE ROC. As seen in section 1, this is one of the overarching aims of the *BSNC*.

The results obtained here do not differ to a great extent from those presented in Bouso and Ruano San Segundo (2021). First, individual changes (Hilpert 2020) in the ROC and the DDC run in parallel across almost the entire 19th century and correlate most significantly in the transition between 1851-1860 ($p < 0.001$). Second, though the aggregate-level division of labour between these two functionally similar constructions seems to be a feature of all individuals of the *BSNC*, their correlation only turned out to be statistically significant for Walter Scott and Charles Dickens's novels ($p < 0.05$). Still, no significant changes in the narrative style of these authors have been identified across their lifespan. In fact, as suggested in section 3.3, Dickens could have been using the ROC consistently in his narrative as another textual device that contributes to his well-known excessiveness and hyperbolic style. Last, the ROC-DDC alternation represents the behaviour of an important part of the novels considered here (82.4%), with only a relatively small group (17.5%) using one of the variants (i.e. the more 'conservative' DDC, which is the older and less expressive variant).

The ROC-DDC alternation seems to have therefore existed with different degrees of entrenchment (Schmid & Mantlik 2015) as a cognitive reality in the minds of the eleven authors in the *BSNC* corpus: as previously shown in Figure 5 all authors are alternating. It remains to be explored what other factors, apart from extravagance, motivated the use of the ROC and propelled its development over the course of the 19th century. After all, as argued by Bouso and Ruano San Segundo, the DDC cannot be treated as 'the single source construction of the ROC' (2021). In this light, to complete the LModE ROC puzzle, it is on the agenda of one of the authors of the paper to explore other semantically and / or formally related constructions (apart from the DDC) that may have modelled and shaped the English ROC over the course of the

LModE period. A closer look at the novels included in the *BSNC* would allow her to see whether the newly coined notion of 'constructional contamination' (Hilpert & Flach Forthcoming; Pijpops, De Smet & Van de Velde 2018; Pijpops & Van de Velde 2016; Torrent 2015) can also be applied to the history of the LModE ROC. The questions to be explored are (i) to what extent would it be possible to claim that the ROC qualifies as a case of multiple source construction (Petré 2012; Van de Velde, De Smet & Ghesquière 2013; Hilpert 2019a, Hilpert 2019b)? and (ii) what other constructions(s) 'bolstered' or strengthened its 19th century mental representation (McColm & Trousdale 2019: 81)?

4. Highlights and Limitations of the BSNC

One of the most salient advantages of the corpus appears to be its size. The *BSNC* is fairly large by comparison with other similar accessible corpora of 19th century fictional narratives, such as CLMET3.0 (15,784,689 words), ARCHER 3.2 (c. 3.3 million words), the *19th Century British Fiction Corpus* (c. 13 million words) or the *19th Century Reference Corpus* (c. 4.5 million words). Another strength of the *BSNC* is that it is freely available to researchers who wish to use it for their own research. They can contact the authors of the paper or download the texts from *Project Gutenberg* themselves. Finally, accessibility also makes the *BSNC* highly manipulable. Thus, the user could add other texts or remove ones we have selected in order to look into their own specific research questions.

As for the disadvantages of the *BSNC*, we can mention the alleged vulnerability of *Project Gutenberg* texts referred to by some scholars, who view the quality of the texts critically because they are created with a system based on contributions from volunteers (see Berglund et al. 2004: 14). Although *Project Gutenberg* texts have proven useful in numerous corpus-linguistic studies, we are aware of potential limitations we may encounter when using texts from this repository. As mentioned by De Smet (2005b), regarding *CLMET*, which was compiled on the basis of *Project Gutenberg* and the *Oxford Text Archive*, 'the corpus had better not be used for the study of phenomena that might lightly attract editorial interventions—for example, matters of punctuation, spelling-related issues, […] or anything that might be seen by an editor as a production error' (2005b: 79). By contrast, for the study of syntactic

structures such as the ROC or the DDC, as demonstrated in the case studies presented here, both corpora seem to be particularly fitting.

Also, our selection is socio-linguistically biased, from a diatopic (British novel), a diastratic (novelists with an upper-class background), and also a diaphasic (sentimental novel) point of view. This clearly hinders the implementation of certain research questions. It is true that, as we have previously mentioned, researchers can expand and reduce the corpus by adding or removing novels, thus adapting the corpus to their own interests (for example, researchers could expand the corpus by adding American sentimental novels). However, we are well aware that our corpus is meant to be representative of a very specific kind of literary English. Finally, a desirable future project will be expanding the sample for the last two sub-periods. As already noted, these are somewhat underrepresented and they do not yet contain texts written by female authors, which would be interesting to have in order to draw gender comparisons for this literary period as well. Last, the *BSNC* is only made available as plain text without part-of-speech (POS) annotation. The latter we plan to implement by means of CLAWS C7 (Garside 1987), whose accuracy rates in other corpora such as *CHELAR* range between 95% and 98.5% (Fanego et al. 2017: 66–69).

Availability

For conditions of use of the *BSNC*, interested researchers can contact the authors of the paper.

Reference line and copyright

British Sentimental Novel Corpus (BSNC). 2019. Compiled by Pablo Ruano San Segundo and Tamara Bouso. Cáceres: Departamento de Filología Inglesa, University of Extremadura.

Appendix

Table 6. Authors and novels in the *BSNC*, per sub-period and exact date of publication

Author	Novel	Words	Year[4]
	***BSNC1* (1798-1820)**		
Jane Austen (1775-1817)	*Sense and Sensibility*	122,986	1811
Jane Austen (1775-1817)	*Pride and Prejudice*	125,216	1813
Walter Scott (1771-1832)	*Waverley*	215,102	1814
Jane Austen (1775-1817)	*Mansfield Park*	163,526	1814
Walter Scott (1771-1832)	*Guy Mannering*	177,491	1815
Jane Austen (1775-1817)	*Emma*	164,057	1815
Walter Scott (1771-1832)	*The Antiquary*	181,519	1816
Walter Scott (1771-1832)	*Rob Roy*	198,455	1817
Walter Scott (1771-1832)	*The Heart of Midlothian*	243,170	1818
Mary Shelley (1797-1851)	*Frankenstein*	78,387	1818
Jane Austen (1775-1817)	*Northanger Abbey*	80,813	1818
Jane Austen (1775-1817)	*Persuasion*	86,690	1818
Walter Scott (1771-1832)	*The Bride of Lammermoor*	130,619	1819
Mary Shelley (1797-1851)	*Mathilda*	51,165	1819
Walter Scott (1771-1832)	*Ivanhoe*	197,860	1820
Walter Scott (1771-1832)	*The Abbot*	187,747	1820
Walter Scott (1771-1832)	*The Monastery*	180,094	1820
	***BSNC2* (1821-1836)**		
Walter Scott (1771-1832)	*Kenilworth*	192,327	1821
Walter Scott (1771-1832)	*Peveril of the Peak*	231,534	1822
Walter Scott (1771-1832)	*The Pirate*	203,671	1822

[4] The year(s) of publication is the year in which a particular novel was published or the years during which the serialised novel was released. For some serialised novels, only one year of publication is given, which means that all instalments of the novel were released in that year. Not all novels included in the corpus were published in a serialised mode, namely those from the Romantic period.

Walter Scott (1771-1832)	*Quentin Durward*	188,567	1823
Walter Scott (1771-1832)	*Redgauntlet*	193,294	1824
Walter Scott (1771-1832)	*St. Ronan's Well*	186,894	1824
Walter Scott (1771-1832)	*The Betrothed*	134,487	1825
Walter Scott (1771-1832)	*The Talisman*	133,178	1825
Walter Scott (1771-1832)	*Woodstock*	201,337	1826
Mary Shelley (1797-1851)	*The Last Man*	179,534	1826
Walter Scott (1771-1832)	*The Fair Maid of Perth*	188,482	1828
***BSNC3* (1837-1850)**			
Charles Dickens (1812-1870)	*Pickwick Papers*	304,448	1836-37
Charles Dickens (1812-1870)	*Nicholas Nickleby*	326,302	1837-39
Charles Dickens (1812-1870)	*Oliver Twist*	159,017	1838-39
William Thackeray (1811-1863)	*Catherine: A Story*	68,506	1839-40
William Thackeray (1811-1863)	*The Paris Sketchbook*	117,191	1840
William Thackeray (1811-1863)	*The History of Samuel Titmarsh*	50,482	1841
Charles Dickens (1812-1870)	*Barnaby Rudge*	256,770	1841
Charles Dickens (1812-1870)	*The Old Curiosity Shop*	218,805	1840-41
William Thackeray (1811-1863)	*The Luck of Barry Lyndon*	12,825	1844
Charles Dickens (1812-1870)	*Martin Chuzzlewit*	340,427	1843-44
Charlotte Brontë (1816-1855)	*Jane Eyre*	187,466	1847
Elizabeth Gaskell (1810-1865)	*Mary Barton*	162,297	1848
William Thackeray (1811-1863)	*The Book of Snobs*	64,778	1848
William Thackeray (1811-1863)	*Vanity Fair*	306,882	1847-48
Charles Dickens (1812-1870)	*Dombey and Son*	358,954	1846-48
Charlotte Brontë (1816-1855)	*Shirley*	217,781	1849
Anthony Trollope (1815-1882)	*La Vendée*	179,506	1850
William Thackeray (1811-1863)	*The History of Pendennis*	359,847	1848-50

Charles Dickens (1812-1870)	*David Copperfield*	359,051	1849-50
	***BSNC4* (1851-1960)**		
William Thackeray (1811-1863)	*Men's Wives*	66,606	1852
William Thackeray (1811-1863)	*The History of Henry Esmond, Esq.*	190,517	1852
Elizabeth Gaskell (1810-1865)	*Cranford*	71,622	1851-53
Elizabeth Gaskell (1810-1865)	*Ruth*	162,993	1853
Charlotte Brontë (1816-1855)	*Vilette*	196,518	1853
Charles Dickens (1812-1870)	*Bleak House*	357,726	1852-53
Charles Dickens (1812-1870)	*Hard Times*	104,049	1854
Elizabeth Gaskell (1810-1865)	*North and South*	184,357	1853-55
William Thackeray (1811-1863)	*The Newcomes*	367,771	1854-55
George Meredith (1828-1909)	*Farina*	34,962	1857
Charlotte Brontë (1816-1855)	*Professor*	89,394	1857
Charles Dickens (1812-1870)	*Little Dorrit*	341,593	1855-57
George Eliot (1819-1880)	*Adam Bede*	217,086	1859
Anthony Trollope (1815-1882)	*The Betrams*	22,527	1859
George Meredith (1828-1909)	*The Ordeal of Richard Feverel*	170,734	1859
William Thackeray (1811-1863)	*The Virginians*	358,865	1857-59
Charles Dickens (1812-1870)	*A Tale of Two Cities*	137,379	1859
George Eliot (1819-1880)	*The Mill on the Floss*	209,306	1860
	***BSNC5* (1861-1870)**		
George Meredith (1828-1909)	*Evan Harrington*	177,464	1861
George Eliot (1819-1880)	*Silas Marner*	72,126	1861
Charles Dickens (1812-1870)	*Great Expectations*	186,545	1860-61
Anthony Trollope (1815-1882)	*Orley Farm*	311,995	1861-62
George Eliot (1819-1880)	*Romola*	229,121	1862-63
William Thackeray (1811-1863)	*Roundabout Papers*	122,723	1863

Elizabeth Gaskell (1810-1865)	*Sylvia's Lovers*	192,758	1863
Anthony Trollope (1815-1882)	*The Small House at Allington*	259,788	1862-64
Anthony Trollope (1815-1882)	*Can You Forgive Her*	318,649	1864-65
George Meredith (1828-1909)	*Rhoda Fleming*	153,555	1865
Charles Dickens (1812-1870)	*Our Mutual Friend*	328,632	1864-65
Elizabeth Gaskell (1810-1865)	*Wives and Daughters*	274,158	1865[5]
George Eliot (1819-1880)	*Felix Holt, the Radical*	183,803	1866
Anthony Trollope (1815-1882)	*The Last Chronicle of Barset*	347,897	1866-67
George Meredith (1828-1909)	*Vittoria*	195,973	1867
Anthony Trollope (1815-1882)	*Phineas Finn*	264,507	1867-69
Anthony Trollope (1815-1882)	*He Knew He Was Right*	344,807	1868-69
Charles Dickens (1812-1870)	*The Mystery of Edwin Drood*	94,960	1870
***BSNC6* (1871-1880)**			
Thomas Hardy (1840-1928)	*Desperate Remedies*	142,703	1871
George Meredith (1828-1909)	*The Adventures of Harry Richmond*	217,622	1870-71
George Eliot (1819-1880)	*Middlemarch*	320,373	1871-72
Thomas Hardy (1840-1928)	*Under the Greenwood Tree*	58,816	1872
Thomas Hardy (1840-1928)	*A Pair of Blue Eyes*	132,063	1872-73
Anthony Trollope (1815-1882)	*The Eustace Diamonds*	271,375	1871-73
Thomas Hardy (1840-1928)	*Far From the Madding Crowd*	139,519	1874
Anthony Trollope (1815-1882)	*Phineas Redux*	263,575	1873-74
George Meredith (1828-1909)	*Beauchamp's Career*	196,134	1874-75

[5] *Wives and Daughters* was published in the *Cornhill Magazine* between August 1864 and January 1866. The last instalment was completed by Frederick Greenwood, as Elizabeth Gaskell had died on 12 November 1865.

Anthony Trollope (1815-1882)	*The Way We Live Now*	355,218	1874-75
George Eliot (1819-1880)	*Daniel Deronda*	312,993	1876
Thomas Hardy (1840-1928)	*The Hand of Ethelberta*	143,425	1875-76
Anthony Trollope (1815-1882)	*The Prime Minister*	284,488	1875-76
George Meredith (1828-1909)	*The House on the Beach*	32,468	1877
Thomas Hardy (1840-1928)	*The Return of the Native*	143,172	1878
George Meredith (1828-1909)	*The Egoist*	188,901	1879[6]
George Meredith (1828-1909)	*The Tale of the Chole*	2,324	1879
Anthony Trollope (1815-1882)	*The Duke's Children*	225,189	1879-80
Thomas Hardy (1840-1928)	*The Trumpet Major*	115,392	1880
***BSNC7* (1881-1890)**			
Thomas Hardy (1840-1928)	*A Laodicean*	140,062	1880-81
Anthony Trollope (1815-1882)	*Doctor Wortle's School*	78,276	1881[7]
Thomas Hardy (1840-1928)	*Two on a Tower*	95,269	1882
George Meredith (1828-1909)	*Diana of the Crossways*	155,486	1884
Thomas Hardy (1840-1928)	*The Mayor of Casterbridge*	117,642	1886
George Meredith (1828-1909)	*Sandra Belloni*	187,176	1887
Thomas Hardy (1840-1928)	*The Woodlanders*	137,209	1887
***BSNC8* (1891-1900)**			
George Meredith (1828-1909)	*One of Our Conquerors*	162,602	1890-91
Thomas Hardy (1840-1928)	*Tess of D'Urbervilles*	151,169	1891
Thomas Hardy (1840-1928)	*Jude the Obscure*	146,493	1894-95
George Meredith (1828-1909)	*The Amazing Marriage*	157,961	1895

[6] Only two of the thirty instalments were released in 1880, hence many literary critics consider *The Egoist* as an 1879 novel.
[7] Although *Doctor Wortle's School* was released as a book in February 1881, the novel was originally published in parts from May to December 1880.

Thomas Hardy (1840-1928)	*The Well-Beloved*	64,022	1892[8]

References

Altick, Richard D. 1957. *The English common reader: A social history of the mass reading public, 1800-1900*. Chicago: University of Chicago Press.

Arnaud, René. 1998. The development of the progressive in 19th century English: A quantitative survey. *Language Variation and Change* 10: 1231–52. doi: 10.1017/S0954394500001265.

Baldick, Chris. 2001. *The concise Oxford dictionary of literary terms*. Oxford: Oxford University Press.

Berglund, Ylva, Alan Morrison, Rowan Wilson, and Martin Wynne. 2004. *An investigation into free ebooks*. Accessed May 30, 2020. https://ora.ox.ac.uk/objects/uuid:76a9cdeb-b325-438e-9ef5-9643919c28e2.

Bodenheimer, Rosemarie. 1994. *The real life of Mary Ann Evans: George Eliot, her letters and fiction*. Cornell: Cornell University Press.

Bouso, Tamara. 2017. *Muttering contempt and smiling appreciation*: Disentangling the history of the Reaction Object Construction in English. *English Studies* 98(2): 194–215.

Bouso, Tamara. 2020a. The growth of the transitivising reaction object construction. *Constructions and Frames* 12(2): 239–71.

Bouso, Tamara. 2020b. The shaping of the Late Modern English Reaction Object Construction. In *Thresholds and ways forward in English Studies*, edited by Lourdes López Ropero, Sara Prieto García-Cañedo and José Antonio Sánchez Fajardo, 179–91. Alicante: Publicaciones de la Universidad de Alicante.

Bouso, Tamara. 2021. *Changes in argument structure: The transitivizing Reaction Object Construction*. Bern: Peter Lang.

Bouso, Tamara, and Pablo Ruano San Segundo. 2021. Another turn of the screw on the history of the reaction object construction. *Functions of Language*. 24 pp. Published online: 7 April 2021.

[8] *The Well-Beloved* was serialised in 1892 and published as a volume on 16 March 1897.

Brook, George Leslie. 1970. *The language of Dickens*. London: André Deutsch.

Busse, Beatrix. 2010. *Speech, writing and thought presentation in a corpus of nineteenth-century English narrative fiction*. Bern: University of Bern.

Bybee, Joan L. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.

Caldas-Coulthard, Carmen Rosa. 1987. Reported speech in written narrative texts. In *Discussing discourse*, edited by Malcom Coulthard, 149–67. Birmingham: University of Birmingham.

Carter, Ronald, and John McRae. 1996. *The Penguin guide to English literature: Britain and Ireland*. London: Penguin Books.

Cichosz, Anna. 2018. The origin of English clause-initial quotative inversion. *Journal of Historical Linguistics* 8(3): 318–55. doi: 10.1075/jhl.17002.cic

Davies, Mark. 2010-. *The Corpus of Historical American English: 400 million words, 1810-2009*. Accessed May 30, 2020. https://www.english-corpora.org/coha/.

De la Pena, Víctor, and Evarist Giné. 2012. *Decoupling: From dependence to independence*. New York: Springer Science & Business Media.

Denison, David, and Nuria Yáñez-Bouza (project coordinators). 2012. *A Representative Corpus of Historical English Registers, version 3.2 (ARCHER-3.2)*. Originally compiled under the supervision of Douglas Biber and Edward Finegan at Northern Arizona University and University of Southern California; modified and expanded by subsequent members of a consortium of universities. Department of Linguistics and English Language: The University of Manchester. Accessed May 30, 2020. https://www.projects.alc.manchester.ac.uk/archer/.

De Smet, Hendrik (compiler). 2008. *Corpus of English Novels (CEN)*. Linguistics Department, Katholieke Universiteit Leuven. Accessed May 30, 2020. https://perswww.kuleuven.be/~u0044428/.

De Smet, Hendrik (compiler). 2005a. *Corpus of Late Modern English Texts (CLMET)*. Linguistics Department, Katholieke Universiteit Leuven. Accessed May 30, 2020. https://perswww.kuleuven.be/~u0044428/.

De Smet, Hendrik. 2005b. A corpus of Late Modern English texts. *ICAME Journal* 29: 69–82.

De Smet, Hendrik (compiler). 2006. *Corpus of Late Modern English Texts (extended version) (CLMETEV)*. Linguistics Department, Katholieke Universiteit Leuven. Accessed May 30, 2020. https://perswww.kuleuven.be/~u0044428/.

De Smet, Hendrik, Hans-Jürgen Diller, and Jukka Tyrkkö (compilers). 2013. *Corpus of Late Modern English Texts, version 3.0 (CLMET3.0)*. Linguistics Department, Katholieke Universiteit Leuven. Accessed May 30, 2020. https://perswww.kuleuven.be/~u0044428/.

De Smet, Hendrik, Susanne Flach, Jukka Tyrkkö, and Hans-Jürgen Diller (compilers). 2015. *Corpus of Late Modern English Texts, version 3.1 (CLMET3.1)*. Katholieke Universiteit Leuven, Freie Universität Berlin, University of Tampere, Ruhr-Universität Bochum. Accessed May 30, 2020. https://fedora.clarin-d.uni-saarland.de/clmet/clmet.html.

Fanego, Teresa, Paula Rodríguez-Puente, María José López-Couso, Belén Méndez-Naya, Paloma Núñez-Pertejo, Cristina Blanco-García, and Iván Tamaredo. 2017. The *Corpus of Historical English Law Reports 1535-1999 (CHELAR)*: A resource for analysing the development of English legal discourse. *ICAME Journal* 41: 53–82. doi: 10.1515/icame-2017-0003.

Flach, Susanne. 2013. Das Ego in der Sprache. *SprachLog*, August 12. Accessed May 30, 2020. http://www.sprachlog.de/2013/08/12/googleology-gets/.

Fonteyn, Lauren, and Andrea Nini. 2020. Individuality in syntactic variation: An investigation of the seventeenth-century gerund alternation. *Cognitive Linguistics* (published online ahead of print). doi: 10.1515/cog-2019-0040.

Fonteyn, Lauren. 2017. The aggregate and the individual: thoughts on what non-alternating authors reveal about linguistic alternations: a response to Petré. *English Language & Linguistics* 21(2): 251–62. doi: https://doi.org/10.1017/S1360674317000119.

Garside, Roger. 1987. The CLAWS word-tagging system. In *The computational analysis of English: A corpus-based approach*, edited by Roger Garside, Geoffrey Leech, and Geoffrey Sampson, 30–41. London: Longman.

Golding, Robert. 1985. *Idiolects in Dickens: The major techniques and chronological development*. London: Macmillan.

Gomme, Andor H. 1978. *Dickens*. London: Evans Bros.

Haspelmath, Martin. 1999. Why is grammaticalization irreversible? *Linguistics* 37(6): 1043–68. doi: 10.1515/ling.37.6.1043.

Head, Dominic (Ed.). 2006. *The Cambridge guide to literature in English* (3rd ed.). Cambridge: Cambridge University Press.

Hilpert, Martin, and Florent Perek. 2015. Meaning change in a petri dish: constructions, semantic vector spaces, and motion charts. *Linguistics Vanguard* 1(1): 339–50. doi: 10.1515/lingvan-2015-0013.

Hilpert, Martin, and Susanne Flach. Forthcoming. A case of constructional contamination in English: Modified noun phrases influence adverb placement in the passive. In *Contrast and analogy in language: perspectives from cognitive linguistics*, edited by Marcin Grygiel. Amsterdam: Benjamins.

Hilpert, Martin. 2011. Dynamic visualizations of language change: Motion charts on the basis of bivariate and multivariate data from diachronic corpora. *International Journal of Corpus Linguistics* 16(4): 435–61. doi: 10.1075/ijcl.16.4.01hil.

Hilpert, Martin. 2013. *Constructional change in English: Developments in allomorphy, word formation, and syntax*. Cambridge: Cambridge University Press.

Hilpert, Martin. 2018. Three open questions in diachronic construction grammar. In *Grammaticalization meets construction grammar*, edited by Evié Coussé, Peter Andersson, and Joel Olofsson, 21–39. Amsterdam: John Benjamins. doi: 10.1075/cal.21.c2.

Hilpert, Martin. 2019a. *Construction grammar and its application to English* (2nd ed.). Edinburgh: Edinburgh University Press.

Hilpert, Martin. 2019b. Interview with Martin Hilpert. Conducted by Marcos Luiz Wiedemer, Marcia dos Santos Machado Vieira, and Maria Maura da Conceição Cezario. *Revista Diadorim* 21(2): 30–43. doi: https://doi.org/10.35520/diadorim.2019.v21n2a30979.

Hilpert, Martin. 2020. The great temptation: what diachronic corpus data do and do not reveal about social change. In *Corpora and the changing society: Studies in the evolution of English*, edited by Paula Rautionaho, Arja Nurmi, and Juhani Klemola, 3–27. Amsterdam: John Benjamins. doi: 10.1075/scl.96.01hil.

Hunt, Leigh, Laurence Sterne, John Hawkesworth, Jean-François Marmontel, Samuel Johnson, Voltaire, Henry Brooke, Oliver Goldsmith, and Henry Mackenzie. 1806. *Classic tales, serious and lively: with critical essays on the merits and reputation of the authors*. London: John Hunt & Carew Raynell.

Janda, Richard D., and Joseph D. Brian. 2003. On language, change, and language change – or, of history, linguistics, and historical linguistics. In *The handbook of historical linguistics*, edited by Richard D. Janda, and Joseph D. Brian, 3–180. Oxford: Blackwell Publishing. doi: 10.1002/9780470756393.ch.

Kim, Evgeny, and Roman Klinger. 2009. A survey on sentiment and emotion analysis for computational literary studies. arXiv:1808.03137v2.

Koplenig, Alexander, and Caroline Müller-Spitzer. 2016. Population size predicts lexical diversity, but so does the mean sea level: Why it is important to correctly account for the structure of temporal data. *PloS One* 11(3). doi: 10.1371/journal.pone.0150771.

Labov, William. 1994. *Principles of linguistic change. Volume I. Internal factors*. Oxford: Blackwell.

Levin, Beth. 1993. *English verb classes and alternations. A preliminary investigation*. Chicago: University of Chicago Press.

Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins.

Liberman, Mark. 2013. The culturomic psychology of urbanization. *Language Log*, 18 August 2013. Accessed May 30, 2020. https://languagelog.ldc.upenn.edu/nll/?p=5985.

Mahlberg, Michaela, Peter Stockwell, Johan de Joode, Catherine Smith, and Matthew Brook O'Donnell. 2016. *CLiC* Dickens: novel uses of concordances for the integration of corpus stylistics and cognitive poetics. *Corpora* 11(3): 433–63. doi: 10.3366/cor.2016.0102.

Mahlberg, Michaela. 2013. *Corpus stylistics and Dickens's fiction*. London: Routledge.

Martínez-Vázquez, Montserrat. 2014. Expressive object constructions in English. A corpus based analysis. *Revista Canaria de Estudios Ingleses* 69:171–86.

McColm, Daniel, and Trousdale Graeme. 2019. Whatever happened to *whatever*? In *Categories, constructions and change in English syntax*, edited by Nuria Yáñez-Bouza, Emma Moore, Linda van

Bergen, and Willem B. Hollmann, 81–104. Cambridge: Cambridge University Press. doi:10.1017/9781108303576.004.

Nevalainen, Terttu, Helena Raumolin-Brunberg, and Heikki Mannila. 2011. The diffusion of language change in real time: Progressive and conservative individuals and the time depth of change. *Language Variation and Change* 23(1): 1–43. doi: 10.1017/S0954394510000207.

Ousby, Ian (Ed.). 1988. *The Cambridge guide to literature in English*. Cambridge: Cambridge University Press.

Paroissien, David. 2000. Characterization. In *Oxford reader's companion to Dickens*, edited by Paul Schlicke, 76–82. Oxford: Oxford University Press.

Petré, Peter, Lynn Anthonissen, Sara Budts, Enrique A. Manjavacas, Emma-Louisa Silva, William Standing, and Odile A. Strik. 2019. Early Modern Multiloquent Authors (EMMA): designing a large-scale corpus of individuals' languages. *ICAME Journal* 43(1): 83–122. doi: 10.2478/icame-2019-0004.

Petré, Peter (project leader), Lynn Anthonissen, Sara Budts, Enrique Manjavacas, Emma-Louise Silva, William Standing, and Odile Aurora Oscar Strik (compilers). 2018. *Early Modern Multiloquent Authors, version 1.0 (EMMA)*. Linguistics Department, University of Antwerp. Accessed May 30, 2020. https://www.uantwerpen.be/en/projects/mind-bending-grammars/emma-corpus/.

Petré, Peter. 2012. General productivity: How *become* waxed and *wax* became a copula. *Cognitive Linguistics* 23(1): 27–65. doi: 10.1515/cog-2012-0002.

Petré, Peter. 2017. The extravagant progressive: an experimental corpus study on the history of emphatic [BE V*ing*]. *English Language & Linguistics* 21(2): 227–50. doi: 10.1017/S1360674317000107.

Pijpops, Dirk, and Freek Van de Velde. 2016. Constructional contamination: How does it work and how do we measure it? *Folia Linguistica* 50: 543–81. doi: 10.1515/flin-2016-0020.

Pijpops, Dirk, Isabeau De Smet, and Freek Van de Velde. 2018. Constructional contamination in morphology and syntax: Four case studies. *Constructions and Frames* 10(2): 269–305. doi: 10.1075/cf.00021.pij.

Pykett, Lyn. 1994. *The nineteenth century sensation novel*. Liverpool: Liverpool University Press.

Piper, Andrew, and Richard Jean So. 2015. Quantifying the weepy bestseller: are commercial novels really more sentimental than literary fiction? *The New Republic*, December 18. Accessed May 30, 2020. https://newrepublic.com/article/126123/quantifying-weepy-bestseller.

Rowland, Ann W. 2008. Sentimental fiction. In *The Cambridge companion to fiction in the Romantic period*, edited by Richard Maxwell, and Katie Trumpener, 191–206. Cambridge: Cambridge University Press. doi: 10.1017/CCOL9780521862523.012.

Ruano San Segundo, Pablo. 2016a. "*Análisis estilístico de los verbos de habla como recurso de caracterización en las novelas de Charles Dickens*." PhD diss., Universidad de Extremadura.

Ruano San Segundo, Pablo. 2016b. A corpus-stylistic approach to Dickens' use of speech verbs: Beyond mere reporting. *Language and Literature* 25(2): 113–29. doi: 10.1177/0963947016631859.

Ruano San Segundo, Pablo. 2017. Reporting verbs as a stylistic device in the creation of fictional personalities in literary texts. *Atlantis. Journal of the Spanish Association for Anglo-American Studies* 39(2). 105–24.

Ruano San Segundo, Pablo. 2018a. Charles Dickens's gender-based use of speech verbs: A stylistic analysis. *Gender and Language* 12(2): 192–217. doi: 10.1558/genl.31137.

Ruano San Segundo, Pablo. 2018b. Dickens's hyperbolic style revisited: Verbs that describe sounds made by animals used to report the words of male villains. *Style* 52(4): 475–93. doi: 10.5325/style.52.4.0475.

Šalinović, Ivana. 2014. Women writers of 19th century Britain. *The Journal of Education, Culture, and Society* 5(1): 218–25. doi: 10.15503/jecs20141-218-225.

Sankoff, Gillian, and Hélène Blondeau. 2007. Language change across the lifespan: /r/ in Montreal French. *Language* 83(3): 560–88. doi: 10.1353/lan.2007.0106.

Schmid, Hans Jörg, and Annette Mantlik. 2015. Entrenchment in historical corpora? Reconstructing dead authors' minds from their usage profiles. *Anglia* 133(4): 583–623. doi: 10.1515/ang-2015-0056.

Semino, Elena, and Mick Short. 2004. *Corpus stylistics. Speech, thought and writing presentation in a corpus of English writing*. London: Routledge.

Stefanowitsch, Anatol, and Stefan Th. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2): 209–43.

Torrent, Tiago Timponi. 2015. On the relation between inheritance and change: The constructional convergence and the construction network reconfiguration hypotheses. In *Diachronic construction grammar*, edited by Jòhanna Barðdal, Elena Smirnova, Lotte Sommerer, and Spike Gildea, 173–212. Amsterdam. John Benjamins. doi: 10.1075/cal.18.06tor.

Torres Cacoullos, Rena, and James A. Walker. 2009. The present of the English future: Grammatical variation and collocations in discourse. *Language* 85(2): 321–354. doi: 10.1353/lan.0.0110.

Traugott, Elizabeth Closs, and Graeme Trousdale. 2013. *Constructionalization and constructional changes*. Oxford: Oxford University Press.

Trousdale, Graeme. 2019. Interview with Graeme Trousdale. Conducted by Ivo da Costa do Rosário, *Soletras* 19(37): 10–19.

Van de Velde, Freek, Hendrik De Smet, and Lobke Ghesquière. 2013. On multiple source constructions in language change. *Studies in Language* 37: 473–89. doi: 10.1075/bct.79.

Walker, Brian, Dan McIntyre, Eliott Land, Hazel Price, and Michael Burke (compilers). 2019. *19th Century British Fiction Corpus, version 1 (HUM19UK)*. University of Huddersfield, Utrecht University, and University College Roosevelt, Middelburg. Accessed May 30, 2020. https://www.linguisticsathuddersfield.com/hum19uk-corpus.

Williams, Paul. 2020. *Novel ideas: Writing innovative fiction*. London: Red Globe Press.

Wolk, Cristoph, Joan Bresnan, Annette Rosenbach, and Benedikt Szmrecsanyi. 2013. Dative and genitive variability in Late Modern English: exploring cross-constructional variation and change. *Diachronica* 30(3): 382–419. doi: 10.1075/dia.30.3.04wol.

Yáñez-Bouza, Nuria. 2011a. ARCHER past and present (1990-2010). *ICAME Journal* 35: 205–36.

Yáñez-Bouza, Nuria. 2011b. ARCHER past and present (1990-2011). Poster presented at the *32nd Conference of the International Computer Archive of Modern and Medieval English (ICAME 32)*. Oslo, Norway, June.