

A corpus-based analysis of Post-Auxiliary Ellipsis voice mismatches in Late Modern English¹

Evelyn Gandón-Chapela

Department of Philology, University of Cantabria, Spain

Avda. de los Castros, 52. 39005. Santander. E-mail: evelyn.gandon@unican.es

Evelyn Gandón-Chapela holds a BA and an MA in English Language and Literature and a PhD in English linguistics from the University of Vigo. In her doctoral dissertation (2016) she investigated Post-Auxiliary Ellipsis in Modern English. At present, she holds a position as Assistant Lecturer at the University of Cantabria. Her research focuses on the study of grammatical change in the recent history of the English language. She has developed research stays at the Universities of Ghent (2011), Chicago (2012), Paris Diderot-Paris 7 (in 2013, 2014 and 2015) and University College London (2017). She is a member of competitive research projects funded by the Spanish Ministry of Science, Xunta de Galicia and University of Cantabria.

¹ This work was supported by the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund under Grant numbers FFI2013-44065-P and FPI BES-2010-030869; the Autonomous Government of Galicia under Grant number GPC2014/060; and the Labex Mobility Grant. The information contained in this article is based on part of Section 3.1.1 of Gandón-Chapela (2020).

A corpus-based analysis of Post-Auxiliary Ellipsis voice mismatches in Late Modern English

This paper analyses Post-Auxiliary Ellipsis voice mismatches between the antecedent clause(s) and the ellipsis site(s) in Late Modern English, using the Penn Parsed Corpus of Modern British English (PPCMBE) (1700–1914). This study focuses on two subtypes of Post-Auxiliary Ellipsis, namely VP ellipsis and Pseudogapping. The results show that voice mismatches were possible in Pseudogapping and VP ellipsis in Late Modern English with low frequencies. This fact serves as counterevidence for the claim about the impossibility of finding voice mismatches in Pseudogapping and confirms corpus-based findings for Present-Day English. As for VP ellipsis, corpus-based studies show that voice mismatches are not attested in Present-Day English. Since they occur in Late Modern English with low frequencies, this contrast may be due to the stylistics or register of the corpora analysed.

Keywords: Post-Auxiliary Ellipsis; voice mismatches; Late Modern English

1. Introduction

The present paper investigates Post-Auxiliary Ellipsis voice mismatches between antecedent clause(s) and the ellipsis site(s) in Late Modern English based on data gathered from the Penn Parsed Corpus of Modern British English (1700–1914). More specifically, two subtypes of Post-Auxiliary Ellipsis constructions have been analysed, namely VP ellipsis and Pseudogapping. The main aim will be to analyse the different cases of voice mismatches in Late Modern English qualitatively, thus providing data from a new period of the history of the English language. Importantly, this will allow a comparison of the characteristics of Post-Auxiliary Ellipsis voice mismatches in Late Modern English with those reported in other theoretical and empirical analyses on Present-Day English data (Sag 1976; Warner 1993; Kehler 2000, 2002; Hardt and Rambow 2001; Kertz 2008, 2013; Merchant 2008, 2013; Bos and Spender 2011; Miller 2011, 2014; Miller and Pullum 2014).

In this paper I will first outline the characteristics of Post-Auxiliary Ellipsis (section 2), which will be the focus of this study. In section 3, I will offer a brief overview of the Penn Parsed Corpus of Modern British English, whose texts have served as the data source of this paper, and I will also explain the methodology used. In section 4, I will offer an overview of the studies that have broached the characteristics of voice mismatches in cases of Post-Auxiliary Ellipsis. In Section 5, I will present the data regarding voice (mis)matches between the antecedent clause(s) and the ellipsis site(s) in Late Modern English and establish a comparison with other empirical works on Present-Day English. Section 6 provides the summary and concluding remarks.

2. Post-Auxiliary Ellipsis

This paper focuses on the study of instances of Post-Auxiliary Ellipsis (Sag 1976; Miller and Pullum 2014), a term which refers to those cases in which a Verb Phrase (VP), Determiner Phrase (DP), Adjective Phrase (AP), Prepositional Phrase (PP) or Adverbial Phrase (AdP) is omitted after one of the following licensors (i.e., those elements that license the occurrence of ellipsis): modal auxiliaries, non-modal auxiliaries *be*, *have* and *do*, and the infinitival marker *to* (the latter taken to be a defective non-finite auxiliary verb, see Miller and Pullum 2014). Here are some examples (licensors appear in italics):²

² Struck-out words represent elided material. Strikethrough is merely used as an expository convention in this paper, as no syntactic or semantic analysis is intended.

- (1) A: Could you open the window for me, please?
 B: Yes, of course I *can* [~~open the window for you~~]_{VP}.
- (2) ‘Look, Dr. Briant, I’ve been with you on this from the word go, you know I *have* [~~been with you on this from the word go~~]_{VP}. [Adapted from the British National Corpus].
- (3) She hadn’t contacted him in 2 years and didn’t really want *to* [~~contact him~~]_{VP}.
- (4) Benjamin is really tall but Mary *is* [~~really tall~~]_{AP} too.
- (5) We_i hope that you have enjoyed our party as much as we_i *did* [~~enjoy~~]_{VP} [yours]_{DP}.
- (6) Lady Chiltern: Robert_i is as incapable of doing a foolish thing as he_i *is* [~~incapable~~]_{AP} [of doing a wrong thing]_{PP}. WILDE-1895,54.391.³

Two subtypes of Post-Auxiliary Ellipsis have been the object of analysis in this study: VP ellipsis and Pseudogapping. Examples (1)–(3) illustrate VP ellipsis, where the VPs *open the window for you*, *been with you on this from the word go* and *contact him* have been omitted after licensors *can*, *have* and *to* respectively. Example (4) will also be treated as a subtype of VP ellipsis where non-verbal material (in this case the AP *really tall*) is omitted after the licensor *be*.⁴ Finally, examples (5) and (6) illustrate Pseudogapping, a subtype of Post-Auxiliary Ellipsis that is very similar to VP ellipsis but in this case a contrastive remnant survives the ellipsis after the licensor (e.g., *yours* and *of doing a wrong thing*) (see Levin 1978, 1986; Carlson 2002; Gengel 2007, 2013; Merchant 2008;

³ Example extracted from the Penn Parsed Corpus of Modern British English (Kroch et al. 2010).

⁴ Following Warner (1993: 5–6; 24–25), (copula) *be* has been taken as an auxiliary here. The examples of Post-Auxiliary Ellipsis given by Warner include modals (*will*, *may*, *should*), perfective *have* and copulative *be* (*Mary is happy to eat meat or fish. –Is she? Well Paul never has been, and John certainly won’t be*). In fact, Warner (1993: 24) contends that “there is a reasonable case that at least ‘progressive’ *be*, ‘passive’ *be*, and *be* with attributive and locative complements should be treated as a single item.”

Miller 2014). Notice that the example of Pseudogapping in (6) also omits non-verbal material (the AP *incapable*) after licensor *be*. However, unlike VP ellipsis, the acceptability of Pseudogapping is more constrained given that it shows a preference for coreferential subjects (Levin 1986: 54; Miller 2014: 76) and comparative contexts (Hoeksema 2006; Sharifzade 2012; Miller 2014), as in (5) and (6), where Pseudogapping takes place in comparative sentences whose subjects corefer with the ones present in the antecedent (*we-we* and *Robert-he*). In addition, contrary to VP ellipsis, Pseudogapping is not licensed by the infinitival marker *to*, as shown in (7) and (8) below:

- (7) It [an enema] leaves some water in you. At least, it seems to * \emptyset me / \emptyset .
- (8) Speaker A: Van Gogh's work is beginning to impress me.
Speaker B: *It's starting to \emptyset me, too / Well! It's finally starting to \emptyset .

[Levin (1986: 54)]

It should be noted that, as Miller and Pullum (2014: 6) contend, the name 'VP ellipsis' is a "very poorly chosen term, [...] because it is neither necessary or sufficient that it should involve ellipsis of a VP". Following Sag (1976: 53), Miller and Pullum (2014: 6) claim that this type of construction should be called 'Post-Auxiliary Ellipsis', since "the defining characteristic is not that a VP is omitted but that a constituent or constituent sequence immediately following an auxiliary is missing" (Miller and Pullum 2014: 6).⁵

⁵ Below, I quote some of the examples provided by Miller and Pullum (2014: 6) that illustrate this point:

- (i) We don't want to cancel the parade, but we could [~~cancel the parade~~]_{VP}. [Post-Auxiliary Ellipsis]
- (ii) You think I'm dumb, but I'm not [~~dumb~~]_{AP}, you know. [Post-Auxiliary Ellipsis]
- (iii) He said there would be results quite soon, and indeed there were [~~results~~]_{NP} [~~quite soon~~]_{AdvP}. [Post-Auxiliary Ellipsis]
- (iv) I couldn't reach him, though I tried [~~to reach him~~]_{VP} several times. [Null Complement Anaphora]

In the present work, the misnomer VP ellipsis has been used in order to differentiate between the cases of ellipsis after an auxiliary (examples (1)–(4)) and those where extra remnants survive ellipsis after the auxiliary(ies), i.e., cases of Pseudogapping (examples (5) and (6)). Following Warner (1993: 5–6), VP ellipsis and Pseudogapping will be treated as two subtypes of Post-Auxiliary Ellipsis that can omit both verbal and non-verbal material.

3. Methodology

The data analysed in the present study were retrieved from the Penn Parsed Corpus of Modern British English (1700–1914) (Kroch et al. 2010), which comprises almost a million words divided into eighteen different genres.⁶ This corpus belongs to the Penn Corpora of Historical English, a collection of running texts and text samples of British English prose from different historical periods (from the Middle English period to the First World War). These texts are available in three different forms: raw text, part-of-speech tagged text and syntactically annotated (parsed) text. Crucially, thanks to its syntactic annotation, it has been possible to search not only for words or word sequences, but also for syntactic structure.

As can be observed, in (i) a VP has indeed been omitted. However, this is not true of examples (ii) and (iii), where what has undergone ellipsis is an AP in the former and an NP together with an AdP in the latter. It should be noted that, as Miller and Pullum (2014: 6) put forward, (iii) “does not even form a constituent according to classical constituency tests; for example, it cannot be clefted”. Finally, (iv) would illustrate a case of Null Complement Anaphora (NCA), a construction where a VP has indeed been omitted, but not after an auxiliary, as in the other cases.

⁶ These are the eighteen genres represented in the Penn Parsed Corpus of Modern British English: Diary, Drama comedy, Non-private letters, Private letters, Biography autobiography, Biography other, Trial proceedings, Sermon, Philosophy, Bible, Educational treatise, Handbook other, History, Law, Fiction, Science medicine, Science other and Travelogue.

Firstly, a manual analysis of twelve raw texts out of 102 files was undertaken in order to retrieve the data (12.24 percent, comprising 112,347 words analysed out of a total of 948,895 words). All of these files belonged to different genres and periods of time of the Penn Parsed Corpus of Modern British English. The examples of Post-Auxiliary Ellipsis found manually were stored in a database, which enabled the analysis of their syntactic patterns in order to draw some generalizations that would allow an automated analysis by means of the java programme CorpusSearch 2. To that end, an algorithm that would retrieve instances of Post-Auxiliary Ellipsis automatically was designed, which achieved a high recall (0.97) at the expense of a lower precision (0.23) (see Gandón-Chapela 2016 for more details on the creation of the algorithm).

4. Voice mismatches in Post-Auxiliary Ellipsis

Since as early as Sag (1976), the possibility of finding voice mismatches between the source and the target of ellipsis has become of interest. Sag (1976) contended that voice mismatches in ellipsis resulted in ungrammatical examples, which led him to propose that syntactic identity needed to hold between the source and the target of ellipsis. However, Dalrymple et al. (1991), Hardt (1993) and Kehler (2000, 2002), among others, provided some examples of ellipsis where voice mismatches were acceptable and served as counterevidence of Sag's (1976) theory:

- (9) A lot of this material *can be presented* in a fairly informal and accessible fashion, and often I do ~~*present it in a fairly informal and accessible fashion*~~. [Dalrymple et al. (1991); originally in Chomsky (1982)]
- (10) The ice cream *should be taken out* of the freezer, if you can ~~*take it out of the freezer*~~. [heard in conversation] [Hardt (1993: 37)]
- (11) This problem *was to have been looked into*, but obviously nobody did ~~*look into the problem*~~. [heard in conversation] [Kehler (2000: 548)]

As may be guessed, the examples of acceptable voice mismatch mentioned in the literature had been found in corpora or corresponded to naturally occurring data.

However, an alternative account of ellipsis in which the latter is licensed by semantics rather than by syntax (Dalrymple et al. 1991, for instance) would predict the acceptability of the previous examples of passive/active mismatch, but also that in (12) below, which is obviously unacceptable. As mentioned in Kertz (2008: 83), ‘canonical’ examples of unacceptable mismatch in ellipsis pattern alike, as in example (12), where a passive antecedent and an intended active target have been paired:

- (12) #The material *was skipped* by the instructors and the TA’s did ~~*skip the material*~~ too.
- (13) The instructors *skipped* the material, and the TA’s did ~~*skip the material*~~ too. [Kertz (2008: 283)]

Notice that (13), where both the antecedent and the target of ellipsis are active, yields a grammatical example of ellipsis. Since both syntactic and semantic approaches about the type of identity that needs to be established between the source and the target of ellipsis do not account for the whole range of data, there emerged other proposals, presented in Kehler (2000, 2002) and Kertz (2008, 2013). Kehler (2000, 2002) attempted to account for the data by resorting to discourse structure instead of syntactic structure. More specifically, Kehler (2000, 2002) contended that examples could be sorted by the type of coherence relation established between the source and the target of ellipsis. In his opinion, those cases of unacceptable voice mismatch are the result of a ‘Resemblance’ coherence relation, where the similarity between the parallel events described in the two clauses is emphasised, as illustrated in example (12) above. When this type of coherence relation is established between two clauses, the two events described are interpreted as independent and receive a parallel reading. In contrast, he argues that the coherence relation established between the two clauses in cases of acceptable voice mismatch would be ‘Cause-Effect’, where what is highlighted is the causal relation between the events reported in the two clauses, as in example (11) above. In this case, the first event is

interpreted as the cause of the second event, that is, there is an asymmetric reading. This leads Kehler to argue that syntactic identity is only required in cases of Resemblance coherence, but not in cases of Cause-Effect coherence. In order to support his theory, he offers several examples of voice mismatches gathered from corpora or spontaneous speech, where the coherence relation is Cause-Effect. Convincingly, he presents these examples together with constructed ones where a Resemblance coherence relation is operative, which demonstrates that acceptability decreases as one changes the type of coherence:

- (14) This problem *was to have been looked into*, but obviously nobody did ~~*look into this problem*~~.
- (15) #This problem *was looked into* by John, and Bob did ~~*look into the problem*~~ too.
- (16) Of course this theory *could be expressed using SDRSs*, but for the sake of simplicity we have chosen not to ~~*express this theory using SDRSs*~~.
- (17) #This theory *was expressed using SDRSs* by Smith, and Jones did ~~*express this theory using SDRSs*~~ too. [Kehler (2000: 548ff.)]

However, Kertz (2008, 2013) makes still another proposal in order to account for the data concerning the acceptability of the voice mismatches presented thus far:

The crucial difference [...] between the cases of acceptable and unacceptable ellipsis examined so far is not syntactic structure, and not discourse structure, but information structure: cases of unacceptable mismatch tend to focus the subject argument of the target clause, while cases of acceptable mismatch focus the auxiliary verb. [Kertz (2008: 284)]

In her view, there exists an informational structural constraint that imposes that contrastive arguments should be preferentially aligned in the syntax, that is, they must appear in the same argument position. This implies that “sensitivity to mismatch follows not from a lack of syntactic parallelism per se, but from a failure to align contrastive arguments, which follows from the argument re-ordering effect of the voice manipulation” (Kertz 2008: 284). Hence, Kertz (2008, 2013) predicts topic continuity

between the antecedent and the ellipsis site in cases of voice mismatches, that is, with the same subject. For that reason, what matters is whether one is dealing with the same agent or not. According to her, then, in instances of acceptable voice mismatches polarity, aspect, tense, mood or modality could be contrastive, as long as there are no contrastive topics.⁷ Here are some examples that illustrate her theory:

- (18) #The material was skipped by the instructors and [the TA's]_{top/foc} did too.
- (19) #The problem was looked into by the committee, just like [the chair]_{top/foc} did.
- (20) A lot of this material can be skipped, and often I [do]_{foc}
[Kertz (2008: 285)]

Kertz (2008: 285) contends that the data presented in Kehler (2000, 2002) “confound information structure and coherence”. On the one hand, those examples of Resemblance coherence provided by Kehler exhibit contrastive topic structures with argument-focus, as illustrated in (21). On the other hand, the examples of Cause-Effect coherence focus the auxiliary, as shown in (22):

- (21) #This problem was looked into by John, and [Bob]_{top/foc} did too.
- (22) This problem was to have been looked into, but obviously nobody [did]_{foc}.

In summary, therefore, Kertz (2008) argues that the auxiliary-focus cases of Post-Auxiliary Ellipsis allow active/passive mismatches whereas the subject-focus cases of

⁷ Kertz's (2008, 2013) explanation of voice mismatches is based on the preservation of the discourse topic both in the ellipsis site and in the antecedent. Since my analysis of the Post-Auxiliary Ellipsis examples relies mostly on the contextual information provided by the passage ranging from the antecedent to the ellipsis site, topic continuity has been checked here by the identification of sentence rather than discourse topics. Given that sentence topics unmarkedly instantiate (fully or partially) discourse topics, my findings can be claimed to corroborate Kertz's postulates.

Post-Auxiliary Ellipsis do not. However, in Kertz (2013) she admitted that her (2008) proposal offered a reanalysis of Kehler's basic generalization without really explaining the source of the effect. In her own words, she had "specifically addressed mismatch, but to explain the data required the stipulation of a constraint enforcing syntactic parallelism for contrastive arguments, which, as was demonstrated in that article, is too strong" (Kertz 2013: 397). Crucially, as shown in the following extract, in Kertz (2013) she proposed the existence of a formal constraint that imposes topic/comment parallelism for all contrastive topic relations as the source of the strong sensitivity to mismatch, be they in ellipsis or nonellipsis contexts:

Specifically, I argue that those cases of ellipsis that show a strong sensitivity to mismatch are identified by the occurrence of a contrastive topic focus in the target clause. In addition to this descriptive analysis, I posit a formal constraint that enforces topic/comment parallelism for discourse segments involved in a contrastive topic relation, as the source of the observed effects. That constraint, crucially, is not specific to ellipsis structures but instead holds for all contrastive topic relations. The analysis, as such, proposes a unified treatment of mismatch violations in both ellipsis and nonellipsis contexts. [Kertz (2013: 391)]

Another important claim in the literature about voice mismatches can be found in Merchant (2008, 2013), who defended that while VP ellipsis permits voice mismatches between the antecedent clause(s) and the ellipsis site, Pseudogapping does not, as illustrated in the following example:

- (23) * Klimt is admired by Abby more than anyone does ~~admire~~ Klee.
(Merchant [2008: 169–170])

As will be shown in the following section, Kertz's (2008, 2013) analysis of Post-Auxiliary Ellipsis voice mismatches does not only account for the data from the Present-Day English period, but also for those from Late Modern English. This is due to the fact

that the data found show that the auxiliary-focus cases of Post-Auxiliary Ellipsis license active/passive mismatches while the subject-focus cases of Post-Auxiliary Ellipsis do not.

Having revised the relevant literature concerning voice mismatches in ellipsis, we are in a position to analyse the examples of voice mismatches attested in our corpus study and check the validity of the predictions of the previous theories in Late Modern English.

5. Data analysis

This section offers a qualitative analysis of the types of voice in Post-Auxiliary Ellipsis examples from Late Modern English in order to check the most common patterns found as well as the impact of theories like the ones presented in Sag (1976), Kehler (2000, 2002), Merchant (2008) and Kertz (2008, 2013). To my knowledge, this aspect of Post-Auxiliary Ellipsis constructions has so far been understudied empirically, and certainly, it has not been studied at all in Late Modern English.

Firstly, in Table 1 it can be observed that the active voice predominates in the target of ellipsis in Pseudogapping, since it comprises almost 90% of the examples found. What is more, in over 88% of the cases, the voice of both the source and the target of ellipsis is active, which means that only around 11% of the instances have passive voice in the target of ellipsis in this type of Post-Auxiliary Ellipsis. The VP ellipsis data, presented in Table 2, point towards the same direction since the active voice predominates in the target of ellipsis (93.76%), although the percentage is slightly higher with respect to Pseudogapping. In addition, the most common pattern attested in VP ellipsis regarding voice is one where both the source and the target of ellipsis are active, yielding over 94% of the total. Thus, the percentage of passive voice in the target of ellipsis in VP ellipsis only represents 5.76% of the examples, which is slightly lower than in Pseudogapping. In sum, the data reveal that active voice predominates in the target of both VP ellipsis and

Pseudogapping (93.76 and 89.53 percent, respectively). If one compares the ratio of active targets and the proportion of active versus passive sentences in the whole corpus (82.66 and 17,34 percent, respectively),⁸ one will reach the conclusion that active voice in the targets of VP ellipsis is a highly significant determinant ($X^2(1) = 89.2956$, $p < .0001$), whereas the statistical significance of active voice targets in Pseudogapping is not so relevant ($X^2(1) = 2.83573$, $p = 0.0921$).

5.1. Pseudogapping

Regarding the mismatches in voice found for cases of Post-Auxiliary Ellipsis, Table 3 displays the data concerning voice mismatches attested in Pseudogapping in Late Modern English. As can be observed, only one voice mismatch (passive-active) has been attested in Pseudogapping in the period under study, which represents 1.16% of the total and is reproduced in (24):

- (24) and dip it in the spawn of Frogs, *beaten* as you would ~~*beat*~~ the whites of eggs, several times letting it dry on pewter, or earthen plates, between each time of dipping it. ALBIN-1736,4.75.

These findings are in line with Miller's study (2014: 87), as he found at least 10 occurrences of voice mismatches in Pseudogapping (out of 1,415, yielding 0.70% of the total) in a large corpus study on Pseudogapping, some of which are quoted below:

⁸ The raw totals of active clauses in PPCMBE have been obtained via the frequency of IP-* segments. On the other hand, the raw frequency of passive clauses has been determined by counting the number of DAN (passive participle *done*), HAN (passive participle *had*) and VAN (other passive participles) tags.

- (25) A whole poached wild striped bass *should be taken to the table* as you would ~~take to the table~~ a Thanksgiving turkey or a crown roast of pork, with a twinkle of extravagance.
- (26) I mean for her to *be dressed –and addressed–* as we would ~~dress and address~~ Becky Sharp, or Ophelia, or Elizabeth Bennet, or Mrs. Ramsay, or Mrs. Wilcox, or even Hester Prynne.

Notice, however, that Miller (2014) only takes into account cases of Pseudogapping with NP remnants in his work, whereas this study takes into account a wider variety of remnants. Interestingly enough, it is worth pointing out that the only mismatch attested in Late Modern English is very similar to the typical case found in Miller (2014), where ellipsis is licensed by the modal auxiliary *would* and the remnant that survives the ellipsis is also an NP, as evinced in (25) and (26). In addition, Miller (2014) mentions that all of the examples of voice mismatches attested constituted cases of comparative Pseudogapping. My data also support this claim, as this example also occurs in a comparative construction (*as*-appositive).

In the light of the only example of voice mismatch found in Pseudogapping, the validity of Kertz's (2008, 2013) theory has also been addressed. Kertz's (2008, 2013) predictions work: our example shares the same patient in both clauses (*the whites of eggs*) and the implicit agent is the same, as there is no *by*-phrase.

In conclusion, on the one hand, our data contradict Merchant's (2013) claim about the impossibility of finding voice mismatches in Pseudogapping. On the other hand, our data confirm Miller's (2014) findings about the possibility of finding examples of voice mismatch in Pseudogapping and Kertz's (2008, 2013) theory about the interaction between information structure and voice mismatches.

5.2. VP Ellipsis

The data concerning voice mismatches in cases of VP ellipsis in Late Modern English are presented in Table 4. Results show that 6 mismatches in voice have been attested in VP ellipsis, yielding 0.58% of the total. It should be noted that the most frequent pattern for voice mismatches is one where the antecedent is active and the target of ellipsis passive (0.48%), as illustrated below:

- (27) I wish heartily, said Wyatt, it was in my power *to entertain* your honour as you ought to be *entertained*. REEVE-1777,7.151.
- (28) A. By Appearance we *engaged* as close as any Ship could be *engaged*, without being aboard of one another. HOLMES-TRIAL-1749,59.1054.
- (29) A. They *could not have disabled* the Cornwall so much as she had been *disabled*, if the Lenox had kept her Station a-stern. HOLMES-TRIAL-1749,59.1054.
- (30) With the other Damage in my Rigging, whereby it became impracticable *to manage* the Ship so readily as otherwise might be *managed*. HOLMES-LETTERS-1749,86.61.⁹
- (31) How happy shall I be can I but *finish* the head this week -as it ought to be *finished*. HAYDON-1808,1,34.886.

Neither Hardt and Rambow (2001) nor Bos and Spenader (2011) found any voice mismatches in their corpus-based studies on VP ellipsis in Present-Day English. On the one hand, Hardt and Rambow (2001) analysed Sections 5 and 6 from the Penn Treebank and they also created a balanced corpus containing a balanced number of positive and negative examples of VP ellipsis (that is, examples in which VPs were or were not elided respectively) by adding positive examples but no negative ones. Regarding Sections 5 and 6 of the Penn Treebank, they reported 15 cases in which the antecedent and the target

⁹ Based on the antecedent, the target of ellipsis of this example could be interpreted as *impracticable to manage* and not *managed*. In that case, this would not constitute a case of voice mismatch between the antecedent and the ellipsis site. However, since both interpretations are available, I have decided not to exclude this example from the classification of voice mismatches.

of ellipsis were both active. In the case of the balanced corpus, they found 97 cases where the antecedent and the ellipsis site were both active and 4 where they were both passive.

On the other hand, Bos and Spenader (2011) reported that they found no voice mismatches in the 486 cases of VP ellipsis mentioned in their corpus study (a sample of the *Wall Street Journal*). A possible explanation for the fact that these voice mismatches in VP ellipsis were found in Late Modern English and not attested in Present-Day English in the studies cited may have to do with the stylistics or register of the corpora under study. Certainly, textual variety constitutes an issue for further research as a factor that may have an impact on the occurrence of VP ellipsis voice mismatches.¹⁰

The results described in Bos and Spenader's (2011) work led Miller (2014: 87) to conclude in his study of Pseudogapping that "it seems at first sight that voice mismatches are actually more frequent with Pseudogapping than with VP ellipsis", admitting at the same time that "this conclusion requires significant qualification". Importantly, he observed that, "contrary to VP ellipsis, mismatches are only found in comparative Pseudogapping. With noncomparative Pseudogapping they are systematically degraded."

We agree with Miller and Pullum (2014: 12) that Kertz (2008) makes the very significant point that the distinction mentioned above is relevant for the acceptability of voice mismatches in Post-Auxiliary Ellipsis: auxiliary focus Post-Auxiliary Ellipsis allows active/passive mismatches whereas subject focus cases of Post-Auxiliary Ellipsis do not.

A close look at examples (27)–(31) reveals that they conform to Kertz's predictions. As can be observed, they comply with the condition of topic continuity

¹⁰ I would like to acknowledge the reviewer who pointed out that this could be the cause of the contrast between the results for Present-Day English and Late Modern English VP ellipsis voice mismatches.

between the antecedent and the ellipsis site, as there is never a *by*-phrase which may contrast with the subject of the active clause. It should be noted, however, that in (28) the subject *we*, in order to refer to *our Ship*, would be a particular subcase of *any Ship*, and therefore there is topic continuity too. What is more, in all of the examples there is a mismatch in modality between the antecedent and the ellipsis site and one contrast in polarity (see (29)).

The validity of Kehler's (2000, 2002) theory about voice mismatches has also been checked with respect to the data attested in our corpus study. According to his hypotheses about discourse relations, there must exist an asymmetric kind of discourse relation between the antecedent and the ellipsis site (temporal succession, concessives, counter-expectation, causality, etc.) in order for voice mismatches to be judged acceptable in ellipsis (a Cause-Effect coherence relation).

In our data there are only two cases where an asymmetric kind of coherence relation holds that may fit his theory, illustrated in (30) and (31). One should be cautious about example (31), however, as it may receive two interpretations: one where there is manner subordination and the reading would be 'like it should be finished', and another where a comparison is established and it would mean 'as it is true that it should be finished'. The ambiguity evinced in this example makes it difficult to discern the type of discourse relation established. But if it were interpreted as involving a comparison between the two clauses, it would be an instance of a symmetric relation, and therefore it would not confirm his theory. The remaining examples (27)-(29), however, exhibit a symmetric kind of discourse relation (either resemblance, parallelism or contrast), and therefore they serve as counterevidence to his theory.

Finally, as shown in Table 4, only one potential example of passive-active mismatch has been found, which is provided in (32) below:

- (32) And believe me my D. Friend when once things *are got* into the state I fear they will ~~*get*~~, nothing can save the country. GEORGE-1763,200.283.¹¹

As can be observed, in (32) the voice mismatch takes place within a relative clause. Once again, this example corroborates Kertz's (2008, 2013) theory because there is topic continuity, given that the subject argument of the antecedent and that of the ellipsis site coincide (*things* and *they*). In this case, there is tense (*are got* vs. *will*) and modality (none vs. *will*) contrast. However, as may be deduced, it does not confirm Kehler's (2000, 2002) theory, since there is no asymmetric discourse relationship established.

6. Concluding remarks

This study has explored voice mismatches in Pseudogapping and VP ellipsis in Late Modern English. Despite the low frequencies of such mismatches in the corpus, my (scarce) results, on the one hand, provide counterevidence for Merchant's (2008, 2013) claim about the impossibility of finding voice mismatches in cases of Pseudogapping and, on the other hand, confirm Miller's (2014) corpus-based findings for Present-Day English. As for VP ellipsis, neither Hardt and Rambow (2001) nor Bos and Spender (2011) found any voice mismatches in their corpus-based study of VP ellipsis in Present-Day English. That stated, the very few examples of VP ellipsis voice mismatch in Late Modern English may be justified by stylistic and/or textual differences in the corpus sources of the studies under scrutiny.

¹¹ As pointed out by a reviewer, this example may be ambiguous as to whether there is a mismatch in voice between the antecedent and the ellipsis site. This is due to the fact that in 18th century English the perfect of the verb *get* could be constructed either with *be* or *have*. Therefore, it could be the case that this is an instance of voice match with an active antecedent. Given that the overall context of King George III's letter is compatible with both interpretations, it remains an open question what is actually intended in this extract.

This paper suggests that voice mismatches in contexts of ellipsis need further research in modern and contemporary English. The results reported here for Late Modern English, based on the largest parsed corpus of this period, have allowed comparison with data from Present-Day English. This investigation has shown that, whereas my data has not given support to Kehler's (2000, 2002) claim that voice mismatches are acceptable in VP ellipsis if there exists an asymmetric discourse relation between the antecedent and the ellipsis site —temporal succession, concessives, etc.—, the validity of Kertz's (2008, 2013) theory, according to which voice mismatches are acceptable as long as there is topic continuity, remains intact.

References

- Bos, Johan, and Jennifer Spender. 2011. "An Annotated Corpus for the Analysis of VP Ellipsis." *Language Resources and Evaluation* 45 (4): 463–494.
- Carlson, Katy. 2002. *Parallelism and Prosody in the Processing of Ellipsis Sentences*. New York: Routledge.
- Dalrymple, Mary, Stuart M. Shieber, and Fernando C. N. Pereira. 1991. "Ellipsis and Higher-Order Unification." *Linguistics and Philosophy* 14: 399–452.
- Gandón-Chapela, Evelyn. 2016. "A Corpus-Based Analysis of Post-Auxiliary Ellipsis in Modern English: Methodological and Theoretical Issues." PhD dissertation, University of Vigo.
- Gandón-Chapela, Evelyn. 2020. *On Invisible Language in Modern English: A Corpusbased Approach to Ellipsis*. London: Bloomsbury Academic.
- Gengel, Kirsten. 2007. "Focus and Ellipsis. A Generative Analysis of Pseudogapping and other Elliptical Structures." PhD dissertation, University of Stuttgart.
- Gengel, Kirsten. 2013. *Pseudogapping and Ellipsis*. Oxford: Oxford University Press.
- Hardt, Daniel. 1993. "Verb Phrase Ellipsis: Form, Meaning, and Processing." PhD dissertation, University of Pennsylvania.
- Hardt, Daniel, and Owen Rambow. 2001. "Generation of VP ellipsis: A Corpus-Based Approach." In *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics, Toulouse, France, 9–11 July 2001*, 290–297.
- Hoeksema, Jack. 2006. "Pseudogapping: Its syntactic analysis and cumulative effects on acceptability." *Research on Language and Computation* 4: 335–352.
- Kehler, Andrew. 2000. "Coherence and the Resolution of Ellipsis." *Linguistics and Philosophy* 23: 533–575.
- Kehler, Andrew. 2002. *Coherence, Reference and the Theory of Grammar*. Stanford: CSLI Publications.
- Kertz, Laura. 2008. "Focus Structure and Acceptability in Verb Phrase Ellipsis." In *Proceedings of the 27th West Coast Conference on Formal Linguistics*, edited by Natasha Abner and Jason Bishop, 283–291. Somerville, MA: Cascadilla Proceedings Project.

- Kertz, Laura. 2013. "Verb Phrase Ellipsis: The View from Information Structure." *Language* 89: 390–428.
- Kroch, Anthony, Beatrice Santorini, and Ariel Diertani. 2010. Penn Parsed Corpus of Modern British English. Accessed March 2016. <http://www.ling.upenn.edu/hist-corpora/PPCMBE-RELEASE-1/index.html>.
- Levin, Nancy S. 1978. "Some identity-of-sense deletions puzzle me. Do they you?" In *Papers from the 14th Regional Meeting of the Chicago Linguistic Society (CLS)*, edited by Donka Farkas, Wesley Jacobsen and Karol Todrys, 229–240. Chicago, IL: University of Chicago, Chicago Linguistic Society.
- Levin, Nancy S. 1986. *Main Verb Ellipsis in Spoken English*. New York: Garland.
- Merchant, Jason. 2008. "An Asymmetry in Voice Mismatches in VP-Ellipsis and Pseudogapping." *Linguistic Inquiry* 39 (1): 169–179.
- Merchant, Jason. 2013. "Voice and Ellipsis." *Linguistic Inquiry* 44 (1): 77–108.
- Miller, Philip. 2011. "The Choice between Verbal Anaphors in Discourse." In *Anaphora Processing and Applications: 8th Discourse Anaphora and Anaphor Resolution Colloquium (DAARC 2011), Lecture Notes in Artificial Intelligence*, vol. 7099, edited by Iris Hendrickx, Sobha Lalitha Devi, António Branco and Ruslan Mitkov, 82–95. Berlin: Springer.
- Miller, Philip. 2014. "A Corpus Study of Pseudogapping and its Theoretical Consequences." In *Empirical issues in syntax and semantics* 10, edited by Christopher Piñón, 73–90. <http://www.cssp.cnrs.fr/eiss10/>.
- Miller, Philip, and Geoffrey K. Pullum. 2014. "Exophoric VP Ellipsis." In *The core and the periphery: Data-driven perspectives on syntax inspired by Ivan A. Sag*, edited by Philip Hofmeister and Elisabeth Norcliffe, 5–32. Stanford, CA: CSLI Publications.
- Sag, Ivan A. 1976. "Deletion and Logical Form." PhD dissertation, MIT.
- Sharifzadeh, Saghie. 2012. "Recherches sur Do, lexique et grammaire." PhD dissertation, University of Paris-Sorbonne.
- Warner, Anthony. 1993. *English Auxiliaries: Structure and History*. Cambridge: Cambridge University Press.

VOICE SOURCE	VOICE TARGET					
	Active		Passive		TOTAL	
	Abs. Freq.	%	Abs. Freq.	%	Abs. Freq.	%
<i>Active</i>	76	98.70	0	0.00	76	88.37
<i>Passive</i>	1	1.30	9	100.00	10	11.63
TOTAL	77	89.53	9	10.47	86	100.00

Table 1. Voice of the source and the target of ellipsis in Pseudogapping

VOICE SOURCE	VOICE TARGET					
	Active		Passive		TOTAL	
	Abs. Freq.	%	Abs. Freq.	%	Abs. Freq.	%
<i>Active</i>	961	99.90	5	7.94	966	94.24
<i>Passive</i>	1	0.10	58	92.06	59	5.76
TOTAL	962	93.76	63	6.14	1,025	100.00

Table 2. Voice of the source and the target of ellipsis in VP ellipsis

VOICE SOURCE	VOICE TARGET	
	Active	
	Abs. Freq.	%
<i>Active</i>	0	0.00
<i>Passive</i>	1	1.16
TOTAL	1	1.16

Table 3. Voice mismatches between the source and the target of ellipsis in Pseudogapping

VOICE SOURCE	VOICE TARGET					
	Active		Passive		TOTAL	
	Abs. Freq.	%	Abs. Freq.	%	Abs. Freq.	%
<i>Active</i>	0	0.00	5	0.48	5	0.48
<i>Passive</i>	1	0.10	0	0.00	1	0.10
TOTAL	1	0.10	5	0.48	6	0.58

Table 4. Voice mismatches between the source and the target of ellipsis in VP ellipsis