

## Research Article

Thomas Hoffmann\*, Thomas Brunner, and Jakob Horsch

# English comparative correlative constructions: A usage-based account

<https://doi.org/10.1515/opli-2020-0012>

received August 26, 2019; accepted March 18, 2020

**Abstract:** English Comparative Correlatives (CCs) consist of two clauses, C1 and C2:

[The more we get together,]<sub>C1</sub> [the happier we'll be.]<sub>C2</sub>

Recently, large corpus studies based on the Corpus of Contemporary American English have unearthed various meso-constructions in English CCs using covarying–collexeme analysis. The present study tests these findings against data from the British National Corpus (BNC), aiming to replicate previous results against data from another standard variety of English (British English) and a corpus that is sampled from a wider range of registers. Over 2,000 CC tokens from the BNC were analyzed with regard to hypotactic features, filler types encountered as comparative elements, and deletion phenomena. Moreover, in contrast to earlier corpus studies (such as Hoffmann, Thomas, Jakob Horsch, and Thomas Brunner. 2019. “The more data, the better: a usage-based account of the English comparative correlative construction.” *Cognitive Linguistics* 30(1): 1–36), the present study also investigates the frequency of the semantically related C2C1 construction (*You will be the happier*<sub>C2</sub>, *the more we get together*<sub>C1</sub>) that previously has been found to be considerably less frequent than its counterpart. The results of the present analysis confirm that English CCs possess more paratactic than hypotactic features and, supporting most of the findings of Hoffmann, Horsch, and Brunner (2019) provide even stronger evidence for the existence of several symmetric meso-constructions.

## 1 Introduction

Belonging to the group of “filler–gap constructions” (Sag 2010), the comparative correlative ([CC] Culicover and Jackendoff 1999) is a construction that in its most basic form consists of two clauses that in the following will be referred to as C1 and C2:

(1) [*The [more]*<sub>FILLER-C1</sub> *we get together.*]<sub>C1</sub> [*the [happier]*<sub>FILLER-C2</sub> *we'll be.*]<sub>C2</sub>

Also known as “comparative conditional construction” (McCawley 1988), “covariational conditional” (Goldberg 2003), “proportional correlative” (den Dikken 2005), “*the*-clauses” (Sag 2010), or “*the... the...* construction” (Cappelle 2011), CCs have attracted substantial interest<sup>1</sup> in the last two decades due to several interesting semantic and syntactic properties they exhibit.

<sup>1</sup> As is evident from the considerable number of introspective studies (Fillmore et al. 1988; McCawley 1988; den Dikken 2003; den Dikken 2005; Borsley 2004; Sag 2010; Cappelle 2011) as well as corpus studies (Hoffmann 2014a, 2014b; Hoffmann 2017a, 2017b; Hoffmann 2018; Hoffmann 2019; Hoffmann et al. 2019).

\* **Corresponding author: Thomas Hoffmann**, Catholic University of Eichstätt-Ingolstadt, Universitätsallee 1, 85072, Eichstätt, Germany, e-mail: thomas.hoffmann@ku.de

**Thomas Brunner, Jakob Horsch:** Catholic University of Eichstätt-Ingolstadt, Universitätsallee 1, 85072, Eichstätt, Germany

Concerning their semantics, CCs are characterized by encoding both asymmetric and symmetric relationships. On the one hand, there is a cause–effect relationship where C1 acts as an independent variable, or “protasis”, and C2 is the corresponding dependent variable or “apodosis” (Goldberg 2003: 220). This becomes apparent when paraphrasing (1): getting together is the cause of us becoming happier. On the other hand, there is parallel, simultaneous change in C1 and C2 over a period of time, adding a layer of symmetric semantics to the CC. Again paraphrasing (1), we can say that we are getting together and becoming happier over a similar time period. Accordingly, the semantics of CCs have also been described as a “pair of semantic differentials” with a “monotonic relationship” (Hoffmann 2014a: 169; Sag 2010: 525–26).

Syntactically, CCs also exhibit interesting properties: both C1 and C2 are introduced by clause-initial elements that resemble the English definite article *the*. We speak of *resemble* here because *the* in CCs does not function as the determiner of noun phrases (NPs), but, instead of a noun, is followed by comparative phrases (e.g., *the more* and *the happier* in (1); cf. Goldberg 2003: 220), which may in turn be followed by optional clauses, as exemplified by *get together* and *we’ll be* in (1). *The* as encountered in CCs has therefore been described as a degree word (den Dikken 2005: 514; cf. the paraphrase *the degree to which we get together more is the degree to which we would be happier*, in which *the* has been replaced by *the degree to which*). Unless construction-specific constraints are added, however, this would mean that other degree words, such as *too* should be able to appear in the same position, which is not the case (cf. \**too the more we get together, too the happier we’ll be.*). Consequently, others suggested treating *the* as construction-specific “fixed substantive,<sup>2</sup> phonologically-specified material” (Hoffmann et al. 2019: 2).

As mentioned above, CCs belong to the group of the so-called filler–gap constructions (Sag 2010) and thus share certain properties with constructions such as WH questions and relative clauses; in particular, the fronting of the so-called “fillers” that in normal declarative clauses would be realized post-verbally (Hoffmann 2018: 182). Thus, compared to their position in declarative clauses (cf. *We get together more.* and *We’ll be happier.*), *more* and *happier* are fronted in (1) just like WH items in WH question (*How often did they get together?* and *What will we be?*). In generative approaches, filler–gap phenomena are frequently explained by a single movement operation such as A-bar movement or WH movement (cf. e.g., Chomsky 1981, 1995, 2000, 2001). However, CCs also possess characteristics that distinguish them from other filler–gap constructions (Sag 2010; Hoffmann et al. 2019: 2), such as the absence of a WH element. Sag, therefore, postulated a construction-specific representation of CCs (Sag 2010: 536) in addition to a general filler-head construction that applies to all various filler–gap constructions. In fact, Sag’s constraint-based analysis assumes two independent constructional constraints for CCs: first, a “*the*-clause construction” that licenses (i.e., accounts for) the two clauses C1 and C2, and second, a “comparative correlative construction” (Sag 2010: 537) that combines the two clauses and computes the construction’s overall semantics (i.e., the CC’s asymmetric and symmetric relationships, see above). Importantly, however, in this approach, C1 and C2 are licensed independently from each other (as two instantiations of the *the*-clause construction), which precludes the possibility of any association of syntactic phenomena across C1 and C2 such as the parallel choice of fillers or parallel deletion phenomena (see Hoffmann 2019: 137–42 and below).

A different approach to CCs is taken by Culicover and Jackendoff, who suggested the CC template given in (2) (Culicover and Jackendoff 1999: 567):

(2) [*the* [<sub>comparative phrase 1</sub> (clause)]<sub>C1</sub> [*the* [<sub>comparative phrase 2</sub> (clause)]<sub>C2</sub>

As has been noted before (Hoffmann 2018: 183, 2019: 137–42) and will be argued in this article, Culicover and Jackendoff’s template is better suited to describe the CC construction because, as Hoffmann (2019) and Hoffmann et al. (2019) have previously demonstrated, there is empirical evidence that suggests that in CCs there are cross-clausal associations between C1 and C2 that cannot be explained by A-bar movement or two separately licensed *the*-clause constructions. In contrast to the

<sup>2</sup> Note that in construction grammar terminology, “substantive” refers to concrete, phonologically fixed linguistic material (as opposed to “schematic” and more abstract material).

latter, we advocate a usage-based constructional account that builds on Culicover and Jackendoff's (1999) analysis.

Construction Grammar (cf., Croft 2001; Goldberg 2006; Bybee 2010; Hoffmann and Trousdale 2013) maintains that the basic unit of language is constructions, that is, pairings of FORM (which can include phonological, morphological as well as syntactic information) and MEANING (which can contain semantic, pragmatic as well as discourse-functional information; Croft and Cruse 2004: 268). Both Sag (2010) and Culicover and Jackendoff (1999) offer a constructional analysis of CCs, since their templates combine a FORM pole with a detailed semantic MEANING pole. However, both their analyses adopt a complete inheritance approach (see Hoffmann 2017a: 323–4) in that they only postulate the least number of constructions/constraints necessary to model English CCs. In contrast to this, we endorse a usage-based construction grammar approach to English CCs (Hoffmann 2014a, 2014b, 2018, 2019, Hoffmann *et al.* 2019). Usage-based approaches (Bybee 2006, 2010) assume that the mental grammar of speakers is “shaped by the repeated exposure to specific utterances” (Hoffmann 2018: 184). In other words, if a pairing of FORM and MEANING (i.e., construction) is encountered by a speaker frequently enough, it will become stored, or entrenched (cf. Croft and Cruse 2004: 276–8), even if it could be licensed by more abstract constructions (i.e., generalized, abstract patterns such as (2) that can be used to produce a great number of instances of a construction). Moreover, usage-based analyses emphasize the role that authentic data play for the input for speakers' generalizations (see also Croft 2001; Barðdal 2008, 2011; Hoffmann 2019: 9–16). As Croft (2001) and Barðdal (2008) note, the input that speakers are exposed to does not always automatically lead to maximally abstract mental generalizations but can also lead to only partly schematic and partly substantive generalizations.

Furthermore, following mainstream usage-based approaches, we assume that mental representations are stored in taxonomic networks (cf. Croft and Cruse 2004: 262–5; Goldberg 2006: 215): speakers first of all encounter specific, substantive instances of a construction (*the more money we come across, the more problems we see*; notorious B.I.G. – *Mo Money Mo Problems*), which are stored in an exemplar-based fashion. Only structures with a high type frequency, that is, those that have been encountered with many different lexicalizations (*the more Bill earned, the more he spent on clothes/the more Jane laughed, the more he felt uncomfortable/the more they heard, the more they wanted to know,...*), all of which share a common meaning, contribute to the entrenchment of a more abstract CC construction such as (2) (cf. Goldberg 2006: 39, 98–101; see also Bybee 1985, 1995; Croft and Cruse 2004: 308–13). Following Hoffmann (2019: 17–18), we take statistically significant frequency effects unearthed by the analysis of corpus data as a proxy for the entrenchment of taxonomic networks (see also Bybee 2010: 10; Gries 2013: 97–101 as well as Stefanowitsch and Flach 2017).

In this article, we focus on the internal structure of CCs as well as the various entrenched constructional patterns. A main goal of the present study is to replicate previous studies to assess the validity of their results. Earlier usage-based studies have already speculated what parts of this network might look like, but these mostly relied on small data samples of around 40 tokens (Hoffmann 2014a, 2014b, 2018). Only two studies rely on a larger data sample of over 1,400 C1C2 tokens (Hoffmann 2019, Hoffmann *et al.* 2019), presenting considerable statistical evidence for the existence of several partly substantive and partly schematic CC constructions (the so-called “meso-constructions”; see below). Yet, this is only a single dataset and, as is standard procedure in any science, requires replication to assess the validity of its results. Moreover, while Hoffmann *et al.* (2019) drew on data from the Corpus of Contemporary American English (COCA), the present study replicates their analysis with more than 2,000 tokens from the British National Corpus (BNC) to test whether any variety-specific factors are at work.

Next, we will take a closer look at several syntactic features of English CCs that are particularly relevant from a usage-based perspective (Section 2). Then we will discuss the data and methodology of the present study (Section 3), followed by the results of the corpus study (Section 4) and a usage-based construction grammar analysis (Section 5).

## 2 Syntactic features of English CCs

Due to their semantic and syntactic properties, extensive research has been conducted on CCs (see Fillmore 1987; Fillmore et al. 1988; McCawley 1988; Michaelis 1994; Culicover and Jackendoff 1999; Borsley 2004; den Dikken 2005; Sag 2010; Cappelle 2011; Kim 2011; Hoffmann 2019). In the following, we will focus on five features that lend themselves to a corpus-based analysis (for other features, see Hoffmann 2019: 44–53) and have important implications for the entrenched constructional network of speakers.

The first of these features concerns the order of C1 and C2. Apart from arrangements like (1), where C1 and C2 appear to iconically encode the construction's cause–effect relationship, a C2C1 arrangement, sometimes referred to as CC' (cf. Culicover and Jackendoff 1999: 549; Hoffmann 2017b), is also possible, as illustrated by (3), a variation of (1):

- (3) [*We'll be (the) happier*]<sub>C2</sub> [*the more we get together.*]<sub>C1</sub>

Whereas the “iconic” (i.e., motivated by the cause–effect semantics of the construction) (Hoffmann 2014a: 32, Hoffmann 2017b) C1C2 order formally features two clause-initial elements (i.e., *the*), *the* is not obligatory in C2 and the comparative phrase is placed at the end of the clause in C2C1 structures. Of course, this raises the question as to whether one of the two orders is preferred over the other. As has been pointed out by Hoffmann (2017b, 2018: 186), Hawkins' competence–performance hypothesis (2004) predicts that the C1C2 order should be preferred by speakers because it corresponds to the cause–effect semantics of the construction.

In fact, Hoffmann (2018), drawing on the BROWN corpora family<sup>3</sup> appears to confirm this, with his data revealing a ratio of 37:1 for the C1C2 over the C2C1 order (Hoffmann 2018: 193). Besides, his diachronic study of the competition between C1C2 and C2C1 (Hoffmann 2017b, 2019: 72–95) indicates that the preference for the former, more iconic, structure has existed since the early Middle English period. This effect, however, has not been investigated in any larger corpus. Our present study now allows us to test this claim using a considerably larger database of more than 2,000 CC tokens, more than 16 times as many as in Hoffmann's (2018) study.

While the competition between C1C2 and C2C1 structures is in itself an interesting phenomenon, there are a couple of properties that only affect C1C2 constructions. Next, we will discuss two syntactic features that have been presented in the literature as an indication of a hypotactic relationship between the C1 and C2 clauses in C1C2 constructions, with C2 being the main clause and C1 being a subordinate clause.<sup>4</sup> While diachronically, English CCs were clearly hypotactic in nature (Hoffmann 2014a: 81, 2017b), we will argue that synchronically the structure has become more paratactic in nature in the present-day English (a claim made in, e.g., Culicover and Jackendoff 1999).

The two hypotactic features that we will examine here are optional *that*-complementizers in C1 (4) and the possibility of optional subject–auxiliary inversion (SAI) in C2 clauses (5a). Note that SAI is not possible in the corresponding declarative clauses (5b).

- (4) [*The more [that]<sub>THAT-complementizer</sub> he says.*]<sub>C1</sub> [*the less I wanna say.*]<sub>C2</sub>

- (5) a. [*The more they work.*]<sub>C1</sub> [*the more [I will/will I]<sub>SAI</sub> pay them.*]<sub>C2</sub> b. \**Will I pay them more.*

In the literature, we find differing opinions regarding the grammaticality of these features: while den Dikken states that *that*-complementizers are possible in both C1 and C2 (2005: 502), Culicover and Jackendoff note that they “cannot appear in C2” (1999: 549). Hoffmann claims that in earlier stages of English, there was an “optional *that*-complementizer” in the C1 clause, whereas “colloquial [modern English] apparently licenses an optional *that* in both C1 and C2” (2014a: 96). However, in his COCA data set of 1,409 C1C2 tokens, *that*-complementizers only appeared in less than 2% (= 24/1,409 tokens) of all C1s and less than 1% of all C2s (6/

<sup>3</sup> The BROWN corpus family consists of the BROWN corpus, the Lancaster–Oslo/Bergen corpus (LOB), the Freiburg Brown Corpus of American English corpus (FROWN), and the Freiburg-LOB Corpus of British English corpus (FLOB).

<sup>4</sup> In C2C1 constructions such as (3), C2 is clearly the main clause and does allow for neither SAI in declarative sentences, cf. [*\*Will we be (the) happier*]<sub>C2</sub> [*the more we get together.*] nor a *that*-complementizer in C2 [*\*That we'll be (the) happier*]<sub>C2</sub> [*the more we get together.*].

1,409 tokens; Hoffmann 2019: 125). For this reason, the present study investigates whether the current data set confirms the low frequency of this phenomenon.

With regard to SAI, Culicover and Jackendoff state that it may occur “marginally [...] in C2 but not C1” (1999: 559) and Hoffmann claims that it is “optional” in C2 but “disfavoured” (2014a: 94). Similarly, den Dikken acknowledges the possibility of SAI in C2 but also states that it is “profoundly ungrammatical” in C1 (2003: 2). In his COCA study, Hoffmann did only find SAI in C2 and again his data confirm that it is disfavoured in American English (with only 3% = 10/337 BE tokens exhibiting SAI; Hoffmann 2019: 127).

As we will show, while historical remnants (see Hoffmann 2014a: 30–9 for a diachronic overview of the development of *that*-complementizers and SAI since Old English) of these two hypotactic features may still be encountered in the present-day English C1C2s, they also appear with extremely low frequencies, leading us to claim that they are no longer central properties of the construction. In fact, there is substantial other evidence for the present-day English C1C2s being largely symmetric structures. A first hint is the identical clause-initial elements in C1 and C2, but much more importantly, as previous research has revealed (Hoffmann 2019, Hoffmann et al. 2019) and the present study will further confirm, there is concrete empirical evidence for an iconic tendency of formal symmetry between C1 and C2 in C1C2 CCs. This leads us to the next two phenomena that the present study investigates: filler types and deletion/truncation phenomena in C1C2 CCs.<sup>5</sup>

There are various filler types that can be inserted into the comparative element slot that follows the clause-initial elements. Apart from adverb phrases (AdvPs) and adjective phrases (AdjPs), as exemplified, respectively, by *more* and *happier* in (1), the comparative element can also be an NP, as in (6), or a prepositional phrase (PP), as in (7):<sup>6</sup>

(6) [*the [more snow]<sub>NP</sub>*]<sub>C1</sub> [*the [less danger]<sub>NP</sub> there is to skiers.*]<sub>C2</sub> (BNC W\_newsp\_brdstht AHC)

(7) [*the [more of them]<sub>PP</sub> you see*]<sub>C1</sub> [*the cheaper it is.*]<sub>C2</sub> (BNC S\_conv KBR)

The filler type that occupies the comparative element slot is of interest because previous research has shown that speakers prefer certain, parallel cross-clausal associations with regard to filler types in C1 and C2 (Hoffmann 2019; Hoffmann et al. 2019). Statistical analyses revealed that despite the many possible filler-type combinations between C1 and C2, it is only symmetric filler types in C1 and C2 such as AdvP<sub>C1</sub>–AdvP<sub>C2</sub>, AdjP<sub>C1</sub>–AdjP<sub>C2</sub>, or NP<sub>C1</sub>–NP<sub>C2</sub> (Hoffmann et al. 2019: 14) that are significantly associated and can, therefore, be considered to have been entrenched as meso-constructions. Again, however, reliable statistical evidence for these patterns only stems from a single, large-scale corpus (Hoffmann 2019; Hoffmann et al. 2019) and requires further empirical corroboration.

Finally, moving on to the optional clause slot of the English C1C2 construction, we will examine deletion and truncation phenomena. While deletion is ungrammatical in normal declarative clauses in the present-day standard English (cf. *\*The price higher./\*The product more interesting.*), the examples (8a–d.) show that the verb BE in both C1 and C2<sup>7</sup> can be optionally left out (see also McCawley 1988; Culicover and Jackendoff 1999; Borsley 2004):<sup>8</sup>

(8) a. *The higher the price is, the more interesting the product is.* b. *The higher the price, the more interesting the product is.* c. *The higher the price is, the more interesting the product.* d. *The higher the*

<sup>5</sup> Again, this variable is not relevant for C2C1 CCs, since the comparative element of their C2 clauses is not a clause-initial filler but appears clause internally; cf. [*We'll be (the) happier*]<sub>C2</sub> [*the more we get together.*].

<sup>6</sup> There is also a so-called “special construction” (SC) (cf. McCawley 1988; den Dikken 2005; Sag 2010) that we will not elaborate on further here due to its very low frequency: [*The [bolder a soldier]<sub>SC</sub> you are,*]<sub>C1</sub> [*the [more (of) a man]<sub>SC</sub> you are.*]<sub>C2</sub>. Hoffmann et al. note that “what is special about these constructions is that pronominal elements (...) do normally not precede indefinite articles” (2019: 5).

<sup>7</sup> Note that while all four possibilities are grammatical and can be encountered in corpora, as Hoffmann et al. point out, there appear to be certain restrictions regarding the deletion of BE (2019: 9; cf. also Culicover and Jackendoff 1999: 554; Borsley 2004: 5), depending on whether the filler is a fronted complement, BE functions as a main verb, the presence of *that*-complementizers, and the interpretation of the clause’s subject as specific or nonspecific.

<sup>8</sup> Again, C2C1 constructions such as (3) do not allow for this type of deletion: cf. [*\*The price higher*]<sub>C2</sub> [*the greater the demand is.*].



*price, the more interesting the product.* (examples from Hoffmann et al. 2019: 7))

In fact, as the following examples found in the BNC demonstrate, we can further distinguish subtypes of BE-deletion in English C1C2s. In addition to full clauses (9), i.e., with the clause slot filled but not with any form of BE, we encountered the retention of BE as a main verb (MV) (10) and an auxiliary verb (11). Similarly, the deleted BE can be a main verb (MV) (12) or an auxiliary (13):

- (9) [*the longer* [**the rain lasted**,]<sub>full clause</sub>]<sub>C1</sub> [*the more quickly* [**the ramparts melted**.]<sub>full clause</sub>]<sub>C2</sub>  
(BNC W\_fict\_prose EFW)
- (10) [*the more successful* [**we are**,]<sub>BE-retained\_mv</sub>]<sub>C1</sub> [*the more* [*we'll attract competition*.]<sub>full clause</sub>]<sub>C2</sub>  
(BNC W\_miscellaneous K9B)
- (11) [*the sooner* [**it is tested**,]<sub>BE-retained\_aux</sub>]<sub>C1</sub> [*the better*.]<sub>C2</sub>  
(BNC W\_non\_ac\_humanities\_arts AR9)
- (12) [*the denser* [**the matter is**,]<sub>BE-deleted\_mv</sub>]<sub>C1</sub> [*the more curvature*.]<sub>C2</sub>  
(BNC W\_fict\_prose FNW)
- (13) [*the more* [**information is requested**,]<sub>BE-deleted\_aux</sub>]<sub>C1</sub> [*the longer it will take (.) to review*.]<sub>C2</sub>  
(BNC W\_ac\_polit\_law\_edu J6N)

In addition to the deletion of BE, C1C2s may also be truncated, i.e., with only the obligatory comparative clause slot filled and no optional clause realized (14). Note that there are also well-known truncated CCs that appear to have become lexicalized (15):

- (14) [*The* [*more data*,]<sub>comparative phrase 1</sub>]<sub>C1</sub> [*the* [*more information*.]<sub>comparative phrase 1</sub>]<sub>C2</sub>  
(BNC W\_commerce FA8)
- (15) The more, the merrier.

Interestingly, confirming earlier results from Hoffmann (2018), Hoffmann et al.'s (2019) COCA corpus study identified significant cross-clausal associations with regard to deletion and truncation phenomena in English C1C2s. As was the case with filler types, these associations reveal a preference for symmetric deletion and truncation: the strongest attraction was determined for the pairs TRUNCATED<sub>C1</sub>–TRUNCATED<sub>C2</sub>, FULL CLAUSE<sub>C1</sub>–FULL CLAUSE<sub>C2</sub>, and BE-DELETED MV<sub>C1</sub>–BE-DELETED MV<sub>C2</sub> (Hoffmann et al. 2019: 21).

The present study thus tries to replicate Hoffmann et al. (2019) as well as Hoffmann (2019) with respect to the just mentioned types of parallel syntactic phenomena in C1 and C2 but also extends it by looking at the competition of C1C2 versus C2C1. We, therefore, aim to give a more detailed account of the constructional network of English CC constructions. In particular, we seek to examine the following features in detail and answer the corresponding questions:

- (1) **C1 and C2 orders:** as has been shown, there are two possible arrangements, C1C2 and C2C1. Do the frequencies in the corpus data confirm an iconic preference for the C1C2 over the C2C1 arrangement? Besides, the C1C2 data will be investigated for the following phenomena (which are not relevant for C2C1 CCs; see above):
- (2) **SAI and *that*-complementizers:** what do the data tell us about the frequency of these phenomena in C1 and C2? Is there evidence for a preference of paratactic over hypotactic features in ModE CCs?
- (3) **Filler types:** various types of syntactic phrases may appear as fillers, including AdvPs, AdjPs, and NPs. Are there cross-clausal associations in the data, as predicted by Hoffmann et al. (2019) and Hoffmann (2019)?
- (4) **Deletion patterns:** similar to filler types, can we determine cross-clausal associations regarding deletion and truncation phenomena?

### 3 Data and methodology

The methodology for the present study largely follows Hoffmann et al.'s study (2019: 9–13), which in turn was based on a number of previous usage-based construction grammar studies (Hoffmann 2014b, 2018).

In contrast to earlier studies, the present article uses corpus data obtained from the BNC to determine the entrenchment of various meso-constructions.

Now, corpus evidence is, of course, not “typically representative of the input [and] output of a particular individual” (Stefanowitsch and Flach 2017: 122). Yet, following Stefanowitsch and Flach’s “corpus-as-output” and “corpus-as-input” hypotheses (2017: 101–3), we assume that corpus data at least afford one window into the mental representations of constructions from a representative sample of language (see also Hoffmann 2019: 17–18). Similar to Hoffmann *et al.* (2019), the data for the present study were extracted from an off-line version of the BNC, which consists of 100 million words and contains samples of both written (about 90%) and spoken (about 10%) language. The off-line version of the BNC does not differ from the online version concerning contents and was chosen because it allows considerably more precise and faster queries using regular expressions. In comparison to COCA, this is only about one fifth the size, but (with the exception of Hoffmann *et al.*’s 2019) still considerably larger than any previous corpus studies of CCs, which were merely based on 1 million word corpora such as the International Corpus of English (ICE) corpus family (Hoffmann 2014b *cf.* also; Hoffmann 2014a, 2018).

The BNC was queried with the following regular expressions (using the CLAWS 5 tag set) to retrieve all instances of CC constructions in the corpus:

- C1C2 patterns:  
“the” [pos = “AJC” | (pos = “AV0” & word = “. + er”) | word = “more|less|worse”] []\* [pos = “AJC” | (pos = “AV0” & word = “. + er”) | word = “more|less|worse”] []\* within s;
- C2C1 patterns:  
[pos = “AJC” | (pos = “AV0” & word = “. + er”) | word = “more|less|worse”] []{0,5} [word = “the”] [pos = “AJC”|(pos = “AV0” & word = “. + er”)|word = “more|less|worse”] within s.

In total, this query yielded 4,256 tokens<sup>9</sup> (3,665 tokens for the C1C2 pattern and 591 tokens for the C2C1 pattern), which were then coded by a team of five student assistants. The student assistants had received intensive training based on a sample data set and were provided with a detailed coding handbook that was composed by the researchers. In addition to this, they attended regular weekly meetings with an author of this study to discuss their progress and any issues they encountered. This author, in turn, checked the student assistants’ work for possible erroneous annotation.

The first task of the student assistants was to discard tokens with false positives, portions of deleted text that were removed for copyright reasons,<sup>10</sup> and the so-called “stacked constructions” where a third “C3” clause follows C1 and C2, as illustrated by (16):<sup>11</sup>

(16) [*the more serious the offence,*]<sub>C1</sub> [*the more difficult to make peace,*]<sub>C2</sub> [*the greater the compensation had to be.*]<sub>C3</sub>

(BNC W\_non\_ac\_soc\_science ADW)

After this task was carried out, a data set with 2,180 relevant C1C2 and C2C1 tokens remained (i.e., 2,076 tokens were discarded). Subsequently, the student assistants coded these tokens for the features that were discussed in the previous section. Table 1 gives an overview of the factors and levels that were coded.

The results of single variables such as ORDER were tested for statistical significance by a chi-square test. Cross-clausal associations, e.g., FILLER TYPE and DELETION, were assessed using a

<sup>9</sup> Note that in contrast to previous studies (Hoffmann 2019, Hoffmann *et al.* 2019), the present study is not just based on a data sample but is based on all tokens that the query yielded.

<sup>10</sup> For copyright reasons, “[e]very 200 words, ten words are removed and are replaced with ‘@’” in the BNC (<https://www.corpusdata.org/limitations.asp>). As noted in Hoffmann *et al.* (2019), however, corpora which have been edited in this manner are “still an indispensable tool for mildly frequent structures such as the CC construction” (2019: 10). Also note that the removed portions should not cause any variation that would significantly affect the results of this study, as the omissions due to copyright do not affect “the ‘relative frequency’ of words, collocates and so on”, since they occur randomly and therefore “without regard to context, affect all words equally the same” (<https://www.corpusdata.org/limitations.asp>).

<sup>11</sup> These structures are, of course, interesting in terms of semantics, as C2 functions both as effect (in relation to C1) and as cause (in relation to C3). They will be analyzed in a separate future study due to the limited scope of this article.

Table 1: Overview of coded variables

Factors	Levels
ORDER	C1C2, C2C1
THAT-COMPLEMENTIZER (for C1 and C2 in C1C2s)	TRUE, FALSE
SUBJECT-AUXILIARY INVERSION (for C1 and C2 in C1C2s)	TRUE, FALSE, NA (if there was no auxiliary verb)
FILLER TYPE (for C1 and C2 in C1C2s)	AdjP, AdvP, NP, PP
DELETION (for C1 and C2 in C1C2s)	Full clause, BE-retained (aux), BE-retained (mV), BE-deleted (aux), BE-deleted (mV), truncated
FILLER-TYPE C1 and DELETION C1 × FILLER-TYPE C2 and DELETION C2	Interaction of the variants for FILLER TYPE and DELETION in C1 vs C2

covarying–collexeme analysis (cf. Stefanowitsch and Gries 2005: 9–11), following Hoffmann et al. (2019). This was done via the *Coll.analysis 3.2a* script for R (Stefanowitsch and Gries 2005: 9; Gries 2007). The *Coll.analysis 3.2a* script uses a Fisher–Yates exact test, which is very precise and handles even small frequencies very well (Gries 2015a: 313). The script provides information on the statistical significance of associations via a value called *collostructional strength*, which is a negative log-transformed  $p$  value (cf. Gries 2007). These have to be interpreted as follows: “values with absolute values exceeding 1.30103 are significant at the level of 5% (since  $10^{-1.30103} = 0.05$ )”. Any value exceeding 2 corresponds to  $p < 0.01$ , and values above 3 indicate a significance level of  $p < 0.001$ . The reason for using negative log-transformed values is the better readability of results located “in the small range of 0.05 and 0”, a range that corresponds to the “most interesting values” (Stefanowitsch and Gries 2005: 7).<sup>12</sup> Note that the covarying–collexeme analysis also provides information as to whether there is repulsion or attraction between two lexemes in a separate column of the output.

In addition to this, the *Coll.analysis 3.2a* script also outputs in separate columns two  $\Delta P$  values, which provide information on the directional dependence of one slot on another. The following example by Gries (2015b) serves to illustrate this: *of course* are two strongly associated lexemes in English, but the preposition *of* co-occurs with many more lexemes than the noun *course*, which is preceded by significantly fewer words. Thus, *course* has a higher cue validity for *of* than the other way round. The  $\Delta P$  value for *course* given *of* ( $\Delta P(\text{course}|\text{of})$ ) is consequently going to be lower than the  $\Delta P$  value for *of* given *course* ( $\Delta P(\text{of}|\text{course})$ ).  $\Delta P$  values range from  $-1$  (strong repulsion) to  $+1$  (strong attraction) and consequently allow for testing to what degree a slot in C1 depends on C2 ( $\Delta P(C1|C2)$ ) as well as the other way round ( $\Delta P(C2|C1)$ ).

Note, as an anonymous reviewer pointed out, that a multivariate analysis of the data that tests the effect of several variables at the same time would, obviously, be preferable to the individual analysis of variables presented below. Yet, as discussed above, none of the clause-internal variables in Table 1 applies to C2C1 CCs. Consequently, these variables can only be investigated in C1C2s (and it is impossible to, e.g., run a mixed effects logistic regression model with these as independent variables and C1C2 vs. C2C1 as the dependent variable). Moreover, even for C1C2s, these variables are not orthogonal but strongly correlated. SAI is only possible if BE is retained (and not deleted). *That*-complementizers are only relevant for full clauses (and irrelevant for truncated clauses). Yet, one interaction that is potentially of interest (and where the variables are not correlated in the ways described above) is the combination of FILLER TYPE × DELETION. We have addressed this issue by collapsing the factors FILLER TYPE C1 and DELETION C1 as well as FILLER TYPE C2 and DELETION C2 and running a covarying–collexeme analysis over these interaction data.

<sup>12</sup> Covarying–collexeme analyses do not automatically correct for multiple testing (Stefanowitsch and Gries 2005: 36, n. 3). However, for each test we indicate which configurations remain significant even after correcting for multiple testing using the conservative Bonferroni method (which divides  $p$  values by the overall number of tests).



## 4 Results

### 4.1 ORDER

First, we present the results for ORDER, i.e., the iconic C1C2 arrangement (17) vs. the C2C1 arrangement (18).

(17) [*the less elaborate you can be,*]<sub>C1</sub> [*the better.*]<sub>C2</sub>

(BNC W\_non\_ac\_humanities\_arts A06)

(18) [*it gets worse*]<sub>C2</sub> [*the longer you look at it.*]<sub>C1</sub> (BNC W\_biography A7C)

Table 2 provides an overview of the frequencies.

**Table 2:** Results for the variable ORDER

ORDER	Tokens
C1C2	2,041
C2C1	139
Total	2,180

As Table 2 shows, these results confirm the strong preference for iconic C1C2 constructions in present-day English as discussed in Section 2, with a ratio of almost 15:1 ( $\chi^2 = 1,659.5$ ,  $df = 1$ ,  $p < 0.001$ ). Yet, while the C2C1 construction has been dispreferred ever since the Middle English period (Hoffmann 2017b, 2019), it is interesting that it still remained a constructional option for speakers, albeit a rather infrequent one. Hoffmann (2018) speculated that this has to do with a pragmatic, focusing function that the C2C1 construction has, as is evident from the distribution of the focus particle *even* (cf. *It becomes **even**<sub>FOCUS</sub> more interesting*<sub>C2</sub>, *the more you think about it.*<sub>C1</sub>. vs? *The more you think about it.*<sub>C1</sub>, *the more interesting it **even**<sub>FOCUS</sub> becomes*<sub>C2</sub>). However, this is a claim that requires further study.

### 4.2 Hypotactic phenomena: *that*-complementizers and SAI

As discussed in Section 2, English C1C2s have been claimed to exhibit syntactic characteristics that suggest a hypotactic relationship between C1 and C2, where C2 is the main clause and C1 the corresponding subordinate clause. Two phenomena that are often cited as evidence for this are optional *that*-complementizers in C1 (19) and optional SAI in C2 (20 and 21). Note that *that*-complementizers have been also claimed to be possible in C2 (22).

(19) [*Now then, the faster [**that**]<sub>THAT-complementizer</sub> we can do this,*]<sub>C1</sub> [*the faster we get on with the game.*]<sub>C2</sub>  
(BNC S\_classroom JA8)

(20) [*the more expensive the decision,*]<sub>C1</sub> [*the more [**will senior management**]<sub>SAI</sub> be involved.*]<sub>C2</sub>  
(BNC W\_commerce G3F)

(21) [*the greater the difference is,*]<sub>C1</sub> [*the less easy [**does it**]<sub>SAI</sub> **become** to dismiss one of the differing parties as a mere inadequate version of the other.*]<sub>C2</sub> (BNC W\_ac\_humanities\_arts ECV)

(22) [*the larger the new settlement becomes (.)*]<sub>C1</sub> [*the less [**that**]<sub>THAT-complementizer</sub> the reduced number of sites you will have available (.)*]<sub>C2</sub>

(BNC S\_pub\_debate HVK)

In the following, we will take a closer look at what the BNC data reveal concerning these phenomena.

### 4.2.1 *That*-complementizers

Tables 3 and 4 provide an overview of the frequencies of *that*-complementizers in the data.

**Table 3:** *That*-complementizers in C1

THAT-complementizer C1	Tokens
TRUE	29
FALSE	2,012
Total	2,041

**Table 4:** *That*-complementizers in C2

THAT-complementizer C2	Tokens
TRUE	2
FALSE	2,039
Total	2,041

First, note that there are significantly more *that*-complementizers in C1 clauses (29 in total) than C2 clauses (only two;  $\chi^2 = 1,926.6$ ,  $df = 1$ ,  $p < 0.001$ ; for an example see (22)). Both instances of *that*-complementizers in C2 are from the spoken part of the corpus, which could be seen as (albeit limited) evidence that if *that*-complementizers appear in C2 at all, they do so in spoken English (Hoffmann 2014a: 96). However, even in C1 clauses, *that*-complementizers are used only marginally: of a total of 2,041 tokens, 29 instances amount to just over 1.42%. This dispreference of *that*-complementizers is again statistically significant ( $\chi^2 = 2,033$ ,  $df = 1$ ,  $p < 0.001$ ). Moreover, this probably also explains why in a covarying–collexeme analysis no pattern emerges as significant across the two clauses (with all four combinations, TRUE–TRUE, TRUE–FALSE, FALSE–TRUE, and FALSE–FALSE having a *collostructional strength* of 0.012; i.e.,  $p > 0.05$ ).

### 4.2.2 SAI

Table 5 presents the frequency of SAI in C2 clauses in the BNC data.

**Table 5:** SAI in C2

SAI C2	Tokens
TRUE	50
FALSE	490
Total	540

As was discussed in Section 2, SAI is commonly cited as evidence for C2 being a main clause. The data do indeed reveal that there is not a single case of SAI in C1. However, of the 540 C2s that contained an auxiliary verb, there were only 50 instances of SAI, amounting to just 9.25%. Again, this effect is strongly significant ( $\chi^2 = 358.52$ ,  $df = 1$ ,  $p < 0.001$ ).

## 4.3 FILLER TYPE

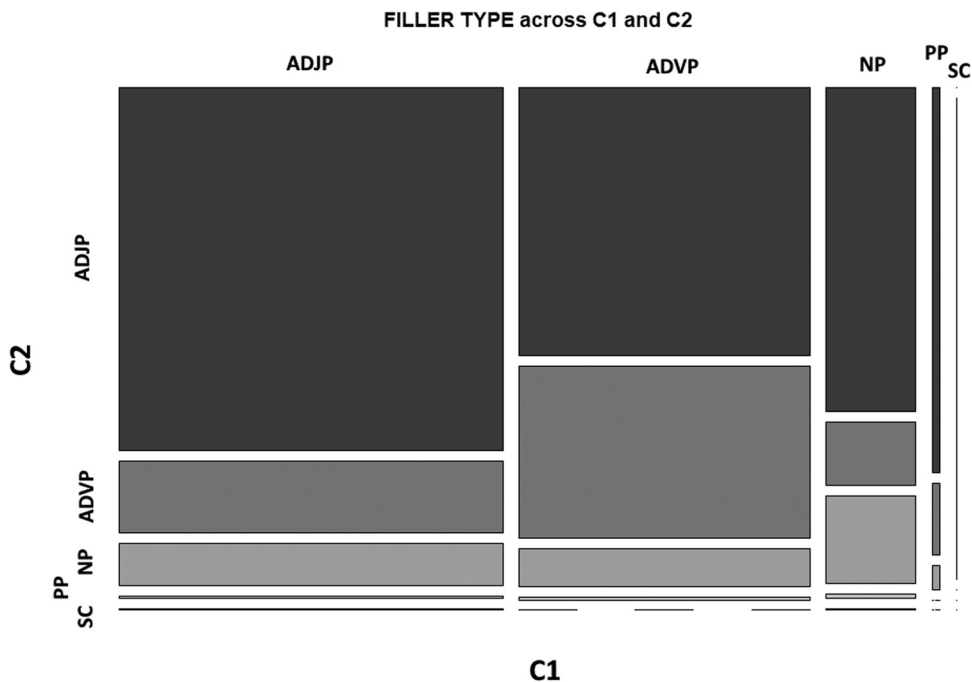
Next, we present the results for the variable FILLER TYPE. Note that here, only data for C1C2 structures were analyzed. Figure 1 gives a first visual impression of the various filler types in C1 and C2; Tables 6 and 7 provide an overview of the frequencies of the various filler types in C1 and C2.

**Table 6:** Filler-type frequencies for C1 in the BNC data

FILLER-TYPE C1	Tokens
AdjP	1,015
AdvP	769
NP	236
PP	20
SC	1
Total	2,041

**Table 7:** Filler-type frequencies for C2 in the BNC data

FILLER-TYPE C2	Tokens
AdjP	1,368
AdvP	463
NP	194
PP	12
SC	4
Total	2,041

**Figure 1:** Filler-type association across C1 and C2 in the BNC data.

The mosaic plot in Figure 1 already suggests a clear tendency toward the mutual association of filler types across C1 and C2: if C1 has an AdjP as a filler, then C2 will highly probably also have an AdjP;<sup>13</sup> the

<sup>13</sup> It is notable that CCs in English clearly prefer AdjPs as filler types, as has been pointed out previously: “AdjPs are by far the most prototypical fillers that speakers encounter in CC constructions” (Hoffmann 2018: 188).

**Table 8:** Results of the covarying–collexeme analysis of the variable FILLER TYPE across C1 and C2 (expected frequency  $\geq 5$ ; significant results with gray background)<sup>15</sup>

C1	C2	Freq. C1	Freq. C2	Observed C1C2	Expected C1C2	Relation	$\Delta P$ (FILLER1  FILLER2)	$\Delta P$ (FILLER2  FILLER1)	Coll. strength
AdvP	AdvP	769	463	275	174.45	Attraction	0.210	0.281	26.647
AdjP	AdjP	1015	1368	765	680.31	Attraction	0.166	0.188	15.068
NP	NP	236	194	43	22.43	Attraction	0.099	0.117	5.183
PP	AdjP	20	1368	16	13.4	Attraction	0.131	0.006	0.801
NP	AdjP	236	1368	159	158.18	Attraction	0.004	0.002	0.315
AdjP	PP	1015	12	6	5.97	Attraction	0	0.003	0.218
AdvP	AdjP	769	1368	428	515.43	Repulsion	-0.182	-0.194	16.626
AdjP	AdvP	1015	463	153	230.25	Repulsion	-0.151	-0.216	15.839
NP	AdvP	236	463	31	53.54	Repulsion	-0.108	-0.063	4.184
AdvP	NP	769	194	62	73.09	Repulsion	-0.023	-0.063	1.315
AdjP	NP	1015	194	88	96.48	Repulsion	-0.017	-0.048	0.942

same applies to the other filler types. The covarying–collexeme analysis, whose results are provided in Table 8,<sup>14</sup> confirms this presumption.

It is striking that just as in Hoffmann et al.’s COCA corpus study (2019), the three significantly associated combinations in the BNC data are symmetric: AdvP<sub>C1</sub>–AdvP<sub>C2</sub> (23), AdjP<sub>C1</sub>–AdjP<sub>C2</sub> (24), and NP<sub>C1</sub>–NP<sub>C2</sub> (25), with highly significant collostructional strengths. Our study thus offers corroborating evidence for the claim that these filler-type combinations form part of the English CC meso-constructural network as specific meso-constructural constructions (cf. Hoffmann et al. 2019: 26):

(23) [*the [longer]<sub>AdvP</sub> the fighting goes on,*]<sub>C1</sub> [*the [less]<sub>AdvP</sub> the chance of a tolerant democratic system emerging.*]<sub>C2</sub> (BNC W\_newsp\_brdsh\_t\_nat\_report AAT)

(24) [*the [more obligatory]<sub>AdjP</sub> an element is,*]<sub>C1</sub> [*the [less marked]<sub>AdjP</sub> it will be.*]<sub>C2</sub> (BNC W\_ac\_soc\_science FRL)

(25) [*the [more equipment]<sub>NP</sub> we have*]<sub>C1</sub> [*the [more problems]<sub>NP</sub> we have*]<sub>C2</sub> (BNC S\_meeting KRY)

With regard to the  $\Delta P$  values provided by the covarying–collexeme analysis, it is notable that the unidirectional associations in the three significantly associated patterns discussed above range from 0.099 ( $\Delta P$  (NP1|NP2)) to 0.281 ( $\Delta P$  (AdvP2|AdvP1)). This suggests a certain degree of entrenchment of these three symmetric patterns at the meso-constructural level but nonetheless creative variation of all possible patterns, including asymmetric ones. These values are very similar to those determined by Hoffmann et al. in their COCA sample (2019: 15).

#### 4.4 DELETION

Further support for the existence of parallel meso-constructural CC templates is evident from the results of the covarying–collexeme analysis for the variable DELETION. Note that only data for C1C2 structures were analyzed. Again, let us first take a look at the raw frequencies, as provided in Tables 9 and 10, and the corresponding mosaic plot (Figure 2).

As was the case with FILLER TYPE, a first glance at the plot in Figure 2 already suggests symmetry across C1 and C2: if there is, e.g., a FULL CLAUSE in C1, it is very likely that a FULL CLAUSE will also appear in C2. We can confirm this intuition by taking a look at the results of the covarying–collexeme analysis presented in Table 11:

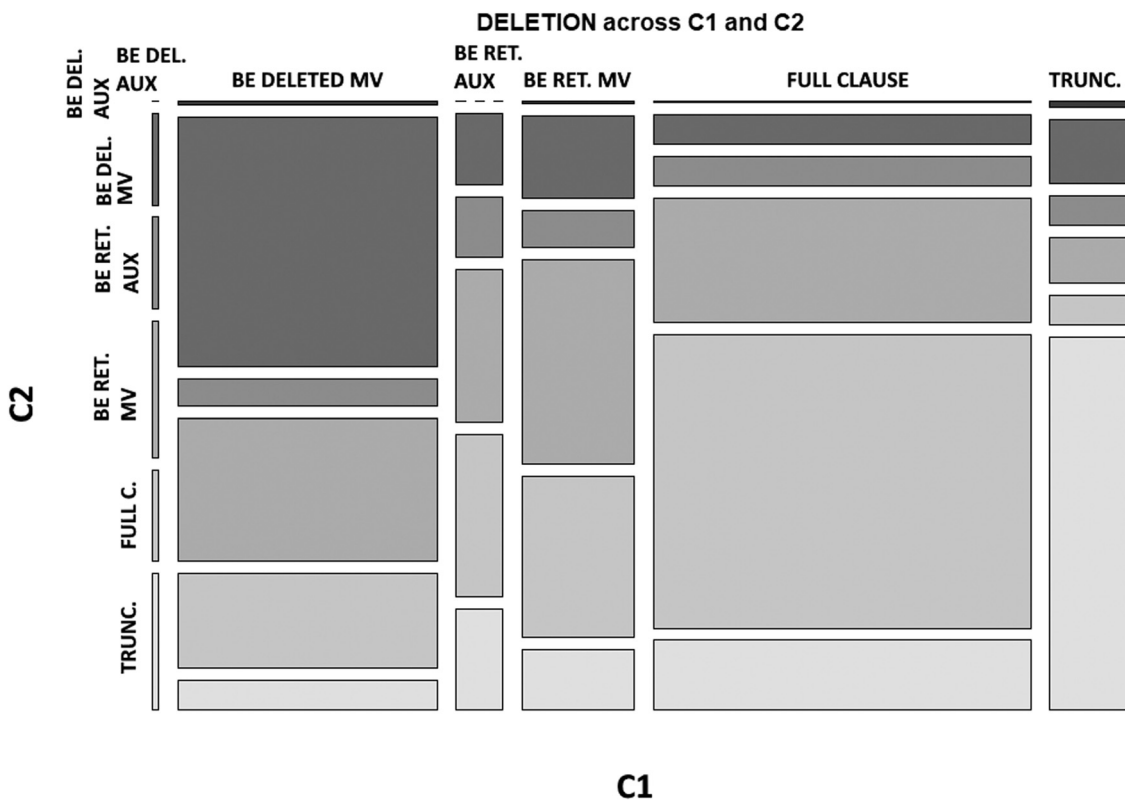
<sup>14</sup> Tables showing the results of covarying–collexeme analyses in this article include only C1C2 combinations with an expected frequency of  $\geq 5$ , a threshold which is commonly “required for goodness-of-fit tests such as the chi-square test” (Hoffmann et al. 2019: 15).

**Table 9:** Deletion frequencies for C1 in the BNC data

BE-DELETION C1	Tokens
FULL_CLAUSE	875
BE-RET. AUX	108
BE-RET. MV	261
BE-DELETED AUX	12
BE-DELETED MV	603
TRUNCATED	182
Total	2,041

**Table 10:** Deletion frequencies for C2 in the BNC data

BE-DELETION C2	Tokens
FULL_CLAUSE	695
BE-RET. AUX	120
BE-RET. MV	500
BE-DELETED AUX	7
BE-DELETED MV	398
TRUNCATED	321
Total	2,041



**Figure 2:** BE-deletion across C1 and C2 in the BNC data.

Similar to the results determined for the variable FILLER TYPE, it is notable that only symmetric combinations exhibit statistically significant attraction, with the strongest one showing up for the



**Table 11:** Results of the covarying–collexeme analysis of the variable DELETION across C1 and C2 (expected frequency  $\geq 5$ , significant results with gray shading)<sup>16</sup>

C1	C2	Freq. C1	Freq. C2	Obs. C1C2	Exp. C1C2	Relation	$\Delta P$ (FILLER1 FILLER2)	$\Delta P$ (FILLER2 FILLER1)	Coll. strength
BE_DELETED_MV	BE_DELETED_MV	603	398	275	117.59	Attraction	0.371	0.491	75.632
TRUNCATED	TRUNCATED	182	321	124	28.62	Attraction	0.575	0.353	64.411
FULL_CLAUSE	FULL_CLAUSE	875	695	469	297.95	Attraction	0.342	0.373	58.365
BE_RETAINED_MV	BE_RETAINED_MV	261	500	97	63.94	Attraction	0.145	0.088	6.177
BE_RETAINED_AUX	BE_RETAINED_AUX	108	120	12	6.35	Attraction	0.055	0.05	1.661
BE_DELETED_MV	BE_RETAINED_MV	603	500	157	147.72	Attraction	0.022	0.025	0.793
BE_RETAINED_AUX	BE_RETAINED_MV	108	500	30	26.46	Attraction	0.035	0.009	0.621
BE_RETAINED_AUX	TRUNCATED	108	321	20	16.99	Attraction	0.029	0.011	0.615
BE_RETAINED_MV	BE_RETAINED_AUX	261	120	18	15.35	Attraction	0.012	0.024	0.576
FULL_CLAUSE	BE_DELETED_MV	875	398	47	170.63	Repulsion	-0.247	-0.386	49.135
BE_DELETED_MV	FULL_CLAUSE	603	695	105	205.33	Repulsion	-0.236	-0.219	25.808
TRUNCATED	FULL_CLAUSE	182	695	10	61.98	Repulsion	-0.314	-0.113	21.175
BE_DELETED_MV	TRUNCATED	603	321	33	94.84	Repulsion	-0.146	-0.229	18.371
TRUNCATED	BE_RETAINED_MV	182	500	15	44.59	Repulsion	-0.178	-0.078	8.428
FULL_CLAUSE	TRUNCATED	875	321	112	137.62	Repulsion	-0.051	-0.095	3.026
TRUNCATED	BE_DELETED_MV	182	398	21	35.49	Repulsion	-0.087	-0.045	2.698
BE_RETAINED_MV	TRUNCATED	261	321	29	41.05	Repulsion	-0.053	-0.045	1.818
BE_RETAINED_MV	BE_DELETED_MV	261	398	39	50.9	Repulsion	-0.052	-0.037	1.586
BE_RETAINED_AUX	BE_DELETED_MV	108	398	14	21.06	Repulsion	-0.069	-0.022	1.336
FULL_CLAUSE	BE_RETAINED_MV	875	500	198	214.36	Repulsion	-0.033	-0.043	1.307
BE_RETAINED_MV	FULL_CLAUSE	261	695	77	88.88	Repulsion	-0.052	-0.026	1.262
BE_DELETED_MV	BE_RETAINED_AUX	603	120	30	35.45	Repulsion	-0.013	-0.048	0.814
BE_RETAINED_AUX	FULL_CLAUSE	108	695	32	36.78	Repulsion	-0.047	-0.01	0.729
FULL_CLAUSE	BE_RETAINED_AUX	875	120	48	51.45	Repulsion	-0.007	-0.031	0.539
TRUNCATED	BE_RETAINED_AUX	182	120	10	10.7	Repulsion	-0.004	-0.006	0.31

BE-DELETED MV<sub>C1</sub>–BE-DELETED MV<sub>C2</sub> pairs (26), for which a very high collostructional strength of 75.632 could be determined. Further significantly attracted pairs are the symmetric TRUNCATED<sub>C1</sub>–TRUNCATED<sub>C2</sub> (27) and FULL CLAUSE<sub>C1</sub>–FULL CLAUSE<sub>C2</sub> (28), which have collostructional strengths of well over 50:

(26) [*the higher the temperature*]<sub>C1</sub> [*the darker the malt.*]<sub>C2</sub> (BNC W\_misc AOA)

(27) [*the more volts*]<sub>C1</sub> [*the more current.*]<sub>C2</sub> (BNC S\_classroom K7F)

(28) [*the further down you press the cap*]<sub>C1</sub> [*the less air enters.*]<sub>C2</sub> (BNC W\_pop\_lore FBN)

The unidirectional cue validities are notably higher than those determined for FILLER TYPES, suggesting a stronger entrenchment. For example, the  $\Delta P$  values for TRUNCATED<sub>C1</sub>–TRUNCATED<sub>C2</sub> are fairly high (0.575 for TRUNCATED1|TRUNCATED2 and 0.353 for TRUNCATED2|TRUNCATED1), which means that TRUNCATION in C1 strongly predicts TRUNCATION in C2 and vice versa. Similarly, high  $\Delta P$  values could be determined for BE-DELETED MV<sub>C1</sub>–BE-DELETED MV<sub>C2</sub> and FULL CLAUSE<sub>C1</sub>–FULL CLAUSE<sub>C2</sub>, with the lowest score being 0.342 (FULL\_CLAUSE1|FULL\_CLAUSE2). The symmetric combinations BE-RETAINED MV<sub>C1</sub>–BE-RETAINED MV<sub>C2</sub> and BE-RETAINED AUX<sub>C1</sub>–BE-RETAINED AUX<sub>C2</sub> exhibit lower, yet still significant collostructional strength values (since values exceeding 1.30103

<sup>15</sup> All in all,  $5 \times 5 = 25$  configurations were tested in Table 9, leading to a Bonferroni-corrected  $p$  value of  $0.05/25 = 0.002$ , which corresponds to a collostructional strength value of 2.70 (since  $\log_{10}(0.002) = -2.69897$ ). Consequently, the three symmetric structures that are positively associated in Table 9 remain significant even after this conservative correction for multiple testing.

correspond to  $p < 0.05$ ). Conversely, significant repulsion could only be determined for asymmetric pairs (see the lower part of Table 11).

#### 4.5 FILLER TYPE × DELETION interaction

As the previous sections showed, both the variables FILLER TYPE and DELETION exhibit significant parallel associations across C1 and C2. While many other variables (such as SAI or *that*-complementizers) are not orthogonal to the other phenomena, FILLER TYPE as well as DELETION should in principle be able to vary independently of each other. At the same time, from a usage-based construction grammar perspective, it is very well possible that associations of these variables can also become entrenched. In order to test this, for each of the two clauses, the levels of the variables FILLER TYPE as well as DELETION were crossed and the resulting complex FILLER TYPE and DELETION factor was subjected to a covarying–collexeme analysis, the results of which can be found in Table 12 on the following page.

As can be seen in Table 12, the statistical analysis reveals 14 significant associations of FILLER TYPE and DELETION across C1 and C2. Similar to the individual results of the two variables, a great number of parallel structures emerge as significant. In fact, 6 of the 14 significant associations have perfectly identical factor combinations (with AdjP\_BE\_DELETED\_MV in C1 and AdjP\_BE\_DELETED\_MV in C2, AdjP\_TRUNCATED in C1 and AdjP\_TRUNCATED in C2 and AdvP\_FULL\_CLAUSE in C1 and AdvP\_FULL\_CLAUSE in C2 being the three most strongly associated patterns with collostructional strength values  $>29$ , i.e.,  $p \ll 0.001$ ). Of the remaining eight, six at least share one feature (either DELETION or FILLER TYPE across C1 and C2). As these results show, neither of these two variables is exclusively associated with a particular feature of the other variable, but the iconic parallel semantics seem to have supported the entrenchment of various parallel structures across C1 and C2.

## 5 Discussion

In the following, we are going to present an analysis of the empirical results that not only sheds more light on the English CC meso-constructional network but also answers important questions concerning the relationship between C1 and C2.

Concerning the order of C1 and C2, the absolute frequencies determined in Section 4.1 reveal a clear preference for the iconic C1C2 over C2C1, with a ratio of 15:1. Note that this is a lower ratio than that determined by a previous BROWN corpus family study by Hoffmann, where it was 37:1 for C1C2 over C2C1 (2019: 193). Nevertheless, we can still speak of a strong tendency toward the iconic C1C2 order. The low frequency of C2C1 structures can thus largely be explained by iconicity.

Next, we turn to the question of whether C1C2s are a hypotactic or paratactic structure in the present-day English. As mentioned in Section 2, previous studies differed in their opinion concerning the grammaticality of hypotactic features in the present-day English CCs. First, there are conflicting views about the possibility of *that*-complementizers in C2 clauses and, second, the status of SAI in C2 has not been conclusively decided, with only vague assertions such that the latter is “marginal” (Culicover and Jackendoff 1999: 559) or “disfavoured” (Hoffmann 2014a: 94). As the absolute frequencies from the BNC data show, we can now confirm that *that*-complementizers, despite their marginal occurrence of just

---

**16**  $6 \times 6 = 36$  configurations were tested in Table 11, corresponding to a Bonferroni-corrected  $p$  value of  $0.05/36 = 0.001388889$ , which translates into a collostructional strength value of 2.86 (since  $\log_{10}(0.001388889) = -2.857332$ ). The only symmetric structure that is positively associated in Table 11 that is no longer significant after this conservative correction for multiple testing is BE-RETAINED AUX<sub>C1</sub>–BE-RETAINED AUX<sub>C2</sub> – all other attraction effects remain significant.

**Table 12:** Results of the covarying–collexeme analysis of the interaction of FILLER TYPE and DELETION across C1 and C2 (only those significant variable pairs are given that have an expected frequency  $\geq 5$ , significant results with gray shading)<sup>17</sup>

C1	C2	Freq. C1	Freq. C2	Obs. C1C2	Exp. C1C2	Relation	$\Delta P$ (FILLER1 FILLER2)	$\Delta P$ (FILLER2 FILLER1)	Coll. strength
AdjP_BE_DELETED_MV	AdjP_BE_DELETED_MV	590	388	264	112.16	Attraction	0.362	0.483	72.189
AdjP_TRUNCATED	AdjP_TRUNCATED	62	296	54	8.99	Attraction	0.749	0.178	38.128
AdvP_FULL_CLAUSE	AdvP_FULL_CLAUSE	612	348	196	104.35	Attraction	0.214	0.318	29.015
AdvP_TRUNCATED	AdvP_TRUNCATED	39	296	30	5.66	Attraction	0.636	0.096	17.934
AdvP_FULL_CLAUSE	AdvP_FULL_CLAUSE	612	230	110	68.97	Attraction	0.096	0.201	8.966
NP_TRUNCATED	AdjP_TRUNCATED	78	296	32	11.31	Attraction	0.276	0.082	8.334
AdjP_BE_RETAINED_MV	AdjP_BE_RETAINED_MV	202	414	70	40.97	Attraction	0.159	0.088	6.472
AdjP_FULL_CLAUSE	AdjP_FULL_CLAUSE	139	230	30	15.66	Attraction	0.111	0.07	3.692
AdvP_BE_RETAINED_AUX	AdjP_TRUNCATED	76	296	20	11.02	Attraction	0.123	0.035	2.368
NP_FULL_CLAUSE	NP_FULL_CLAUSE	116	108	13	6.14	Attraction	0.063	0.067	2.169
AdvP_FULL_CLAUSE	NP_FULL_CLAUSE	612	108	44	32.38	Attraction	0.027	0.114	2.027
NP_FULL_CLAUSE	AdjP_BE_RETAINED_MV	116	414	34	23.53	Attraction	0.096	0.032	1.96
AdjP_BE_DELETED_MV	AdjP_BE_RETAINED_MV	590	414	138	119.68	Attraction	0.044	0.056	1.799
AdjP_FULL_CLAUSE	NP_FULL_CLAUSE	139	108	13	7.36	Attraction	0.044	0.055	1.545

1.42% in the data, are almost exclusively present in C1 clauses, with only two cases of *that*-complementizers having been encountered in C2 clauses, both of which were from the spoken register. This supports Hoffmann's claim of this phenomenon appearing only in "colloquial" English (2014a: 96). With regard to SAI, the frequencies determined in the BNC data again confirm the literature: with a percentage of just under 10 in C2, we can assume that this is indeed a marginal phenomenon.

This, of course, leads to the question why these features still exist in the present-day English. As Hoffmann (2017b, 2019) noted, both features were already disfavored by the end of the Middle English period. One explanation would be to treat these mere historical relics that have survived into the present-day English. Alternatively, Goldberg (building on previous work by Bolinger 1977; Haiman 1980) claims in Corollary A of her Principle of No Synonymy that if two constructions differ syntactically but are semantically synonymous, they must encode some kind of pragmatic difference (1995: 67). Take a look again at (19), repeated below as (29) and the corresponding alternative without a *that*-complementizer in C1:

(29) [*Now then, the faster [that]<sub>THAT-complementizer</sub> we can do this,*]<sub>C1</sub> [*the faster we get on with the game.*]<sub>C2</sub>  
(BNC S\_classroom JA8)

(30) [*Now then, the faster we can do this,*]<sub>C1</sub> [*the faster we get on with the game.*]<sub>C2</sub>

Is it plausible that the difference between (29) and (30) is a pragmatic one, and if so, which one? After all, variable *that*-complementizers can also be found in other constructions such as the N + BE + (*that*) construction (e.g., *The truth is that she was wrong.* vs. *The truth is she was wrong.*, cf. Mantlik and Schmid 2018). Mantlik and Schmid (2018: 191) argue that the N + BE + *that* construction combines "a topicalizing with a focusing function", with the N slot being topicalized and the *that*-clause expressing information that is focused as expressing some fact. While the CC constructions are, of course, completely different types of structures, this explanation could possibly also explain the (limited) use of *that*-complementizers in these constructions: the preposed filler phrases in C1 and C2 can be argued to be topicalized in CC constructions (i.e., they are what the two clauses are "about"). Adding focus particles to C1C2 constructions normally seems unacceptable (cf. *The more you (? even) think about it, the more interesting it (? even) becomes.*). Hoffmann (2018: 186–7) claimed that only C2C1 allow for such focus particles in C2 (cf. *It becomes even more interesting, the more you think about it.*). In light of Mantlik and Schmid's study, however, it might be possible that *that*-complementizers are used in C1C2s to express focused information. Future studies will, however, have to investigate the prosody of these structures to see whether there is any independent evidence for this claim (e.g., a focus accent on *that*).

The same question can also be asked about optional SAI in C2 – does the inverted order of subject and auxiliary serve any pragmatic purposes in these cases? Compare (20), repeated below as (31), with (32):

(31) [*the more expensive the decision,*]<sub>C1</sub> [*the more [will senior management]<sub>SAI</sub> be involved.*]<sub>C2</sub>  
(BNC W\_commerce G3F)

(32) [*the more expensive the decision,*]<sub>C1</sub> [*the more senior management will be involved.*]<sub>C2</sub>

(31) and (32) are semantically synonymous, so Goldberg's Principle of No Synonymy again predicts some kind of pragmatic difference. Again, it is possible that SAI is used in cases where information is focused, but, as mentioned above, future studies will have to seek additional evidence for this (admittedly) bold claim.

Regardless of the reasons for the continued presence of these hypotactic features, their frequency is so low in the present-day English that we can interpret them as evidence for English CCs having a strong tendency toward a more paratactic rather than hypotactic relationship between C1 and C2. This claim

17 176 FILLER TYPE × DELETION configurations were tested in Table 12, corresponding to a Bonferroni-corrected *p* value of  $0.05/176 = 0.0002840909$ , which translates into a collostructional strength value of 3.55 (since  $\log_{10}(0.0002840909) = -3.546543$ ). After using this conservative correction method, the top eight attracted configurations remain significant, five of which have a fully symmetrical combination of FILLER TYPE and DELETION in C1 and C2. The remaining three attracted configurations share at least one other feature.

receives further support by the findings on the variables filler types and BE-deletion/truncation phenomena.

The results of the covarying–collexeme analysis revealed at least three statistically significantly associated filler-type combinations ( $\text{AdvP}_{C1}$ – $\text{AdvP}_{C2}$ ,  $\text{AdjP}_{C1}$ – $\text{AdjP}_{C2}$ , and  $\text{NP}_{C1}$ – $\text{NP}_{C2}$ ) and five statistically significantly associated deletion phenomena combinations (BE-DELETED  $\text{MV}_{C1}$ –BE-DELETED  $\text{MV}_{C2}$ , TRUNCATED $_{C1}$ –TRUNCATED $_{C2}$ , FULL CLAUSE $_{C1}$ –FULL CLAUSE $_{C2}$ , BE-RETAINED  $\text{MV}_{C1}$ –BE-RETAINED  $\text{MV}_{C2}$ , and BE-RETAINED  $\text{AUX}_{C1}$ –BE-RETAINED  $\text{AUX}_{C2}$ ). Moreover, several filler-type and deletion patterns are together significantly associated across C1 and C2. Based on our statistical analysis, these combinations can therefore be considered entrenched as meso-constructions in the English CC network. This, consequently, corroborates the findings of previous research that found exactly the same five meso-constructions based on the data from a different corpus, the COCA (Hoffmann 2019, Hoffmann et al. 2019). What is striking is that all of the statistically significant cross-clausal associations are symmetric, despite the many other combinations that are possible and were indeed encountered in the corpus data (attesting to the productivity of the CC construction). This is, therefore, the clear evidence that supports our claim that the central properties of Modern English CCs are paratactic, not hypotactic.

Finally, the productivity of CCs indicates that we still have to postulate a maximally abstract macro-construction such as (2) to account for all the various observed variable combinations. At the same time, our usage-based approach supports Hoffmann et al.'s view (2019) that, in addition to this, the English taxonomic CC network also contains the above meso-constructions with strong parallel features.

## 6 Conclusion

The present large-scale corpus study has provided new insights into the various phenomena of English CCs. Our analysis confirmed the findings of previous studies but also uncovered new, hitherto unknown, facts about the English CC:

- Concerning the order of C1 and C2, we have been able to show that the iconic C1C2 structure is strongly preferred over C2C1 ones, with a ratio of 15:1. Since focus particles appear to be only acceptable in C2 in C2C1 structures, we assume that C2C1s encode a pragmatic, focusing function.
- Furthermore, the present study investigated *that*-complementizers and SAI. Both of these features can be found in the present-day English CCs, albeit with very low frequencies. This suggests that these two features are no longer central properties of the English CC construction, which appears to be significantly more paratactic in nature, as suggested by the symmetric cross-clausal associations that were determined using statistical analyses. In line with Goldberg's Principle of No Synonymy (1995: 67), we tentatively raised the hypothesis that both these features have a pragmatic function of expressing focus. Yet, these claims clearly require future empirical corroboration.
- Finally, the covarying–collexeme analyses of filler types and deletion phenomena confirm the findings of Hoffmann et al.'s COCA corpus study (2019), i.e., cross-clausal C1C2 associations that are evidence for entrenched meso-constructions. These cross-clausal attraction phenomena could only be found for symmetric structures. Significant repulsion was only found for asymmetric structures.

The implications of the above results are twofold: first, they provide further evidence that CCs are rather paratactic than hypotactic in nature, thus encoding the symmetric semantics of CCs. The use of hypotactic features might be explained with pragmatic functions, but this is something that future studies will have to investigate in more detail.

Second, since they cannot be explained by many previous approaches that treat C1 and C2 as two independent structures that are licensed separately from each other, our data offer further support for Hoffmann et al.'s (2019: 32) assumption that meso-constructural templates play a significant role in the present-day CC construction network.



## References

- Barðdal, Johanna. 2008. *Productivity: Evidence from Case and Argument Structure in Icelandic*, Constructional Approaches to Language 8. Amsterdam: John Benjamins.
- Barðdal, Johanna. 2011. "Lexical vs. structural case: a false dichotomy." *Morphology* 21(1): 619–54.
- Bolinger, Dwight. 1977. *Meaning and Form*. London: Longman.
- Borsley, Robert D. 2004. "An approach to English comparative correlatives." In *Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar, Center for Computational Linguistics, Katholieke Universiteit Leuven*, ed. Stefan Müller, 70–92. Stanford, CA: CSLI Publications.
- Bybee, Joan L. 1985. *Morphology: A Study into the Relation between Meaning and Form*. Amsterdam: John Benjamins.
- Bybee, Joan L. 1995. "Regular morphology and the lexicon." *Language and Cognitive Processes* 10: 425–55.
- Bybee, Joan L. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.
- Cappelle, Bert. 2011. "The *the... the...* construction: meaning and readings." *Journal of Pragmatics* 43(1): 99–117.
- Chomsky, Noam. 1981. *Lectures on Government and Binding*, Studies in Generative Grammar 9. Dordrecht, Netherlands: Foris Publications.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2000. "Minimalist inquiries: the framework." In *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*, ed. Roger Martin, David Michaels, and Juan Uriagereka, 89–155. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2001. "Derivation by phase." In *Ken Hale: A Life in Language*, ed. Michael Kenstowicz, 1–52. Cambridge, MA: MIT Press.
- Croft, William. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.
- Croft, William, and D. Alan Cruse. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- Culicover, Peter W., and Ray Jackendoff. 1999. "The view from the periphery: the English comparative correlative." *Linguistic Inquiry* 30(4): 543–71.
- den Dikken, Marcel. 2003. "Comparative correlatives and verb second." In *Germania et alia: Alinguistic webschrift for Hans den Besten*, ed. Jan Koster, and Henk van Riemsdijk, available at [http://odur.let.rug.nl/\\_koster/DenBesten/DenDikken.pdf](http://odur.let.rug.nl/_koster/DenBesten/DenDikken.pdf)
- den Dikken, Marcel. 2005. "Comparative correlatives comparatively." *Linguistic Inquiry* 36(4): 497–532.
- Fillmore, Charles J. 1987. "Varieties of conditional sentences." *Proceedings of the Eastern States Conference on Linguistics* 3: 163–82.
- Fillmore, Charles J., Paul Kay, and Mary C. O'Connor. 1988. "Regularity and idiomaticity in grammatical constructions: the case of *let alone*." *Language* 64(3): 501–38.
- Goldberg, Adele. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: The University of Chicago Press.
- Goldberg, Adele. 2003. "Constructions: a new theoretical approach to language." *TRENDS in Cognitive Sciences* 7(5): 219–24.
- Goldberg, Adele. 2006. *Constructions at Work*. Oxford: Oxford University Press.
- Gries, Stefan Th. 2007. *Coll. Analysis 3.2a. A Program for R for Windows 2.x*.
- Gries, Stefan Th. 2013. "Data in construction grammar." In *The Oxford Handbook of Construction Grammar*, ed. Thomas Hoffmann, and Graeme Trousdale, 93–108. Oxford: Oxford University Press.
- Gries, Stefan Th. 2015a. "The role of quantitative methods in cognitive linguistics: corpus and experimental data on (relative) frequency and contingency of words and constructions." In *Change of Paradigms – New Paradoxes: Recontextualizing Language and Linguistics*, ed. Jocelyne Daems, Eline Zenner, Kris Heylen, Dirk Speelman, and Hubert Cuyckens, 311–25. Berlin & New York: De Gruyter Mouton.
- Gries, Stefan Th. 2015b. "Quantitative designs and statistical techniques." In *The Cambridge Handbook of English Corpus Linguistics*, ed. Douglas Biber, and Randi Reppen, 50–71. Cambridge: Cambridge University Press.
- Haiman, John. 1980. The iconicity of grammar: isomorphism and motivation. *Language* 56: 515–40.
- Hawkins, John A. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
- Hoffmann, Thomas. 2014a. *Comparing English Comparative Correlatives*. Post-doc thesis, Osnabrück University.
- Hoffmann, Thomas. 2014b. "The cognitive evolution of Englishes: the role of constructions in the dynamic model." In *The Evolution of Englishes: The Dynamic Model and Beyond*, Varieties of English around the World: G49, ed. Magnus Huber, Sarah Buschfeld, Thomas Hoffmann, and Alexander Kautzsch, 160–80. Amsterdam: John Benjamins.
- Hoffmann, Thomas. 2017a. "Construction grammars." In *The Cambridge Handbook of Cognitive Linguistics*, ed. Barbara Dancygier, 310–29. Cambridge: Cambridge University Press.
- Hoffmann, Thomas. 2017b. "Construction grammar as cognitive structuralism: the interaction of constructional networks and processing in the diachronic evolution of English comparative correlatives." *English Language and Linguistics* 21(2): 349–73.
- Hoffmann, Thomas. 2018. "Comparing comparative correlatives: the German vs. English construction network." In *Constructional Approaches to Syntactic Structures in German*, ed. Hans C. Boas, and Alexander Ziem, 181–203. Berlin: Mouton de Gruyter.

- Hoffmann, Thomas. 2019. *English Comparative Correlatives: Diachronic and Synchronic Variation at the Lexicon–Syntax Interface*, Studies in English Language. Cambridge: Cambridge University Press.
- Hoffmann, Thomas, and Graeme Trousdale. 2013. *The Oxford Handbook of Construction Grammar*, Oxford Handbooks in Linguistics. Oxford: Oxford University Press.
- Hoffmann, Thomas, Jakob Horsch, and Thomas Brunner. 2019. “The more data, the better: a usage-based account of the English comparative correlative construction.” *Cognitive Linguistics* 30(1): 1–36.
- Kim, Jong-Bok. 2011. “English comparative correlative construction: interactions between lexicon and constructions.” *Korean Journal of Linguistics* 36(2): 307–36.
- Mantlik, Annette, and Hans-Jörg Schmid. 2018. “That-complementizer omission in N + BE + *that*-clauses – register variation or constructional change?” In *The Noun Phrase in English: Past and Present*, ed. Alex Ho-Cheong Leung and Wim van der Wurff, 187–222. Amsterdam: John Benjamins.
- McCawley, James D. 1988. “The comparative conditional construction in English, German, and Chinese.” *Berkeley Linguistics Society* 14: 176–87.
- Michaelis, Laura A. 1994. “A case of constructional polysemy in Latin.” *Studies in Language* 18: 45–70.
- Sag, Ivan A. 2010. “English filler-gap constructions.” *Language* 86(3): 486–545.
- Stefanowitsch, Anatol, and Susanne Flach. 2017. “The corpus-based perspective on entrenchment.” In *Entrenchment and the Psychology of Language Learning: How we Reorganize and Adapt Linguistic Knowledge*, ed. Hans-Jörg Schmid, 101–27. Berlin: De Gruyter.
- Stefanowitsch, Anatol, and Stefan Th. Gries. 2005. “Covarying collexemes.” *Corpus Linguistics and Linguistic Theory* 1(1): 1–43.