Small corpora and low-frequency phenomena: *try and* beyond contemporary, standard English

Les petits corpus et les phénomènes de basse fréquence: try and au-delà de l'anglais standard et contemporain

Daniel Ross

Electronic version

URL: <http://journals.openedition.org/corpus/3574>

ISSN: 1765-3126

Publisher

Bases ; corpus et langage - UMR 6039

Electronic reference

Daniel Ross, « Small corpora and low-frequency phenomena: *try and* beyond contemporary, standard English », *Corpus* [Online], 18 | 2018, Online since 09 July 2018, connection on 08 September 2020.

URL : <http://journals.openedition.org/corpus/3574>

This text was automatically generated on 8 September 2020.

© Tous droits réservés

Small corpora and low-frequency phenomena: *try and* beyond contemporary, standard English

Les petits corpus et les phénomènes de basse fréquence: try and au-delà de l'anglais standard et contemporain

Daniel Ross

Introduction

- 1 The recent introduction of massive corpora for linguistics research may give the impression that using small corpora is unrepresentative and outdated, but for a researcher interested in looking beyond English and a few other languages, beyond normal written usage by normal adult native speakers, beyond contemporary usage, or in any narrow domain or genre, large corpora are scarce. For the vast majority of linguistic questions we might ask, large corpora are not available. Furthermore, the type of corpus required depends on the research question itself. For some questions, corpora readily lend themselves to efficient answers, even with relatively small datasets. For example, finding the most frequent words, morphological forms or syntactic structures in a language is possible with any corpus that is a well-chosen sample of the language. But other questions are much more challenging to address. Larger and larger corpora may make corpus research sufficient for even many low-frequency phenomena, but comparable research is more difficult when looking at historical development, dialectal variation and language production during acquisition, especially for theoretical questions where microvariation is crucial.
- 2 A syntactician looking at the limits of grammaticality typically will rely on native speaker judgments because corpora do not provide the nuanced data required. Today, massive corpora contain so much data that researchers may sometimes consider unattested usage in a corpus to be a proxy for ungrammaticality, especially with the growing emphasis on using real-data in research rather than lab-generated sentences.

Furthermore, while of course the best approach, and with slightly different results, would be to use both native speaker judgments and consider usage data from corpora, there are certain types of data that are either extremely difficult or impossible to acquire with a grammaticality judgment survey. In those cases, corpora may be a necessary methodological alternative to explore variation in usage from multiple perspectives.

- 3 Specifically, although adult speakers of the standard, contemporary English may be readily available for consultation, adjusting any of those parameters to look at microvariation, or other languages, makes consultation with speakers less feasible. Research on diachronic development may only be based on textual evidence, not speaker intuition, and also may be difficult to interpret because modern speakers only have intuitions for the current (e.g., resulting) grammatical system, so looking for examples of gradual change in historical corpora is challenging. Acquisition data as well presents a challenge for speaker intuitions: children are for practical reasons, such as research ethics guidelines, less accessible to researchers, and their metalinguistic skills are still developing, while second language learners often report what they perceive to be the correct usage patterns as opposed to how they actually speak. Thus corpora are a desirable alternative in these situations. More importantly, the development of such corpora supports an indefinite number of research projects using the same resources as opposed to requiring each researcher to independently gather data from native speakers, which should result in more efficient and productive research in the field.
- 4 At the same time, these types of data are not those that typically correspond to large corpora. Historical corpora are limited in size based on available textual data. Corpora representing dialectal variation are typically much smaller than for the standard language, both due to availability of texts and the interests of the average researcher. Learner corpora are growing in size and popularity but still represent a high level of diversity across different learner populations. And beyond English, as well as a few in-demand major European and other languages, finding large corpora for other languages is unlikely, making consultation with native speakers the only viable research method for questions about many languages. Overall, the availability of large (or massive) corpora is based on the simple laws of supply and demand: are the texts available, and do enough researchers desire access to such a corpus? The result is an overrepresentation of English and a handful of other languages; of standard dialects; of written usage; of contemporary rather than historical texts; and of adults who are native speakers of the language. This bias is only strengthened by the desire to create large corpora that will have the most appeal to a variety of corpus linguists, while small corpora can be more easily made for special purposes. Of course for many reasons large (and massive) corpora are a welcome tool, but there are a number of questions, in fact the majority of potential questions a linguist could ask, that fall outside the types of research such general corpora can facilitate.
- 5 The alternative is to work with small corpora, which make up for their limited size by having specialized types of data and being a more accessible alternative for researchers who need to make their own corpora for special projects. It is not the goal of this article to discuss the terminology of 'small' versus 'large' corpora, but see the other articles in this volume. Instead the current article investigates what it means for small corpora to be *large enough* to address research questions and to comment on the state of the art of

corpus research for low-frequency phenomena beyond contemporary, standard English. In other words, are small corpora a barrier for research on low-frequency phenomena beyond contemporary, standard English, or do they effectively fill the gaps?

- 6 Therefore, this article takes the *try-and-V* construction as a case study for research questions about a low-frequency phenomenon using small corpora for diachronic development, dialectal variation, and acquisition.

1. Try-and-V pseudocoordination

- 7 This article explores the feasibility of investigating low-frequency phenomena using *try-and-V* pseudocoordination as a case study. Specifically, this construction displays an unusual morphosyntactic constraint: neither verb may be inflected. Given its unusual properties we might expect to find variation across dialects, that such constructions would be difficult or slow during acquisition, and that historical usage could elucidate this puzzling synchronic grammatical phenomenon.

- 8 Usage of *try-and-V* is restricted to contexts with bare, uninflected verbs (Carden & Pesetsky 1977; Ross 2014) such as imperatives, infinitives and the present tense (except 3rd sing.), resulting in the *Bare Form Condition* (BFC):¹

- (1) We try and use corpora effectively.
- (2) *He tries and use(s) corpora effectively.
- (3) *We tried and use(d) corpora effectively.

- 9 In fact, the BFC requires a level of morphological analysis (as opposed to phonological form and syntactic features). The second verb is selected as a complement and necessarily a bare infinitive. This is illustrated by the licensing of an infinitive form in the present tense, as in (4):

- (4) We try and be effective corpus linguists. (cf. *We be...)

- 10 The bare form for the first verb is more difficult to explain. Subject-verb agreement is still required in this construction, as shown in (5), which along with (2) results in paradigmatic gaps for the construction:

- (5) *He try and use(s) corpora effectively.

- 11 Thus the first verb, but not the second, is morphosyntactically part of and determined by the matrix clause. Why then, given that only the second verb is a bare infinitive, should inflected usage, of the first verb only, not be permitted?

- (2') *He tries and use corpora effectively.

- 12 An explanation can be derived from a well-known property of other types of pseudocoordination, and coordination more generally: parallelism. For example, *go-and-V* is not restricted to only bare forms, but both verbs must have parallel inflection:

- (6) I went and got a book.
- (7) *I went and have a book.

- 13 Thus, three constraints combine to produce the BFC, and the apparently surface-level restriction emerges from a systematic analysis (Ross 2014, 2015): (A) the first verb participates in normal subject-verb agreement; (B) the second verb is a bare infinitive complement; and (C) the two verbs are required to have parallel inflection. The second verb is not inflected as a bare infinitive, nor can the first be inflected lest it have non-parallel inflection: both verbs must be uninflected, and the BFC emerges.

- 14 This article asks to what extent it is possible to effectively research the BFC beyond contemporary, standard English. Section 1.1 discusses the construction in more detail and provides background information about the construction in standard English as well as some considerations for the feasibility of corpus research with *try-and-V*. Section 2 addresses the diachronic development of the BFC using historical corpora; Section 3 addresses potential dialectal variation and the distribution of the BFC using dialectal corpora; and Section 4 addresses acquisition of the BFC, with a focus on first-language acquisition. These three case studies stretch the available corpora to their limits, but with successful results showing that the BFC is robust.

1.1. Try-and-V and corpora

- 15 The *try-and-V* construction introduces some challenges for corpus research, and conversely is therefore an ideal case study on low-frequency phenomena.
- 16 Although the construction is morphosyntactic in nature, not just idiomatic or lexicalized, corpus research on the *try-and-V* construction is effectively limited to a single verb: *try*. In fact, several other subject control verbs such as *be sure* and *remember* can appear as the first verb in the construction (Ross 2014: 211), but they are too infrequent, especially in written usage, to be thoroughly investigated and statistically analyzed in most corpora. However, we can reasonably investigate the construction through its usage with *try*: the verb *try* is the 127th most frequent word in the *Corpus of Contemporary American English* (COCA: Davies 2008), with about 650 instances per million words, including about 13 instances of *try-and-V* per million words.²
- 17 Furthermore, we must not only show that *try-and-V* is attested in the data, but instead potential *variation* in the BFC and that otherwise violations of the BFC are *not attested*. Given that the BFC represents standard usage in the most pragmatically relevant contexts (imperatives, present tense, etc.), if there are any variation or violations of the BFC, these will be much less frequent than *try-and-V* itself. Substantial data is needed, across all relevant corpora to be consulted.
- 18 But what exactly should be considered *sufficient* regarding these small corpora? Intuitively, we should be able to identify enough clear usage of *try-and-V* that any unattested forms seem unlikely to occur. We can also approach this question quantitatively: we need sufficient data to be able to use statistical tests showing that unattested forms are below a certain threshold for potentially appearing in additional data, assuming that the corpus is representative. Of course the clearest evidence would be that of clear violations of the BFC in a particular dialect, during a certain time period, or during acquisition, but as will be shown in the following sections, the corpus results show the BFC is generally consistent across varieties. Although addressing some aspects of the BFC is challenging, including the need to demonstrate the absence of violations, current small corpora are effective for this case study, and the outlook is optimistic.

2. Diachronic development

- 19 *Try-and-V* is often claimed to be a relatively recent phenomenon or dismissed as a modern error by prescriptivists, but this is an example of a recency illusion. Even some theoretical work on the construction has underestimated its age, such as Rohdenburg

(2003: 237) and Vosberg (2006: 224–234) who date it to the late 1800s based only on examples found in the *Oxford English Dictionary*.³ But *try-and-V* has a nearly 500-year history in English, as shown by Hommerberg & Tottie (2007) and Tottie (2012). The origin of *try-and-V* is investigated in Section 2.1, and its later development in Section 2.2, and the emergence of the modern Bare Form Condition in Section 2.3.

2.1. Origin of *try-and-V*

- 20 Historical text is a relatively well-represented variant of English in corpora. However, there are still some limitations and fundamental problems with this type of research. First, the corpora are limited to written data, and there are either practical or inherent limitations to the amount of text available, especially for earlier time periods. But even if a sufficient corpus is available, the results may be difficult to interpret for speakers of contemporary English, and of course no native speakers of historical English can be consulted. Therefore, generalizations may need to be based on a small number of available examples with sufficient context for making nuanced distinctions. And that is the case with *try-and-V* in the *Early English Books Online* (EEBO TCP) corpus.
- 21 Tottie (2012:210) claims that *try-and-V* predates *try-to-V* based on raw frequencies of the sequences *try and [verb]* and *try to [verb]* in EEBO, but this claim is problematic when the data is manually filtered. EEBO provides many examples of *try-and-V*, but without part of speech tagging this potentially ambiguous construction is challenging. Examples of pseudocoordination are often ambiguous with normal coordination, and crucially it is exactly this ambiguous usage from which the *try-and-V* construction (with semantics like *try-to-V*) developed, so interpreting early examples of *try-and-V* is even more challenging. The raw sequence *try and [verb]* might be normal coordination (*try and fail*), not complementation via pseudocoordination (*try and [=to] win*):
- (8) I will aduventure, or trie and seeke my fortune.
(Baret 1573, cited in Tottie 2012: 207)
- 22 It was from examples like (8), presumably literal coordination ('test and seek out'), that the control-reading emerged due to pragmatic equivalence in some contexts.
- 23 EEBO is not a traditional corpus, but rather a database of Early Modern English books, made available as EEBO TCP with full-text search by the Text Creation Partnership. It is not balanced, represents a mix of genres, and has texts of varying lengths written by different authors. Additionally, the corpus extends from 1475 through 1700, but Ross (2013) consulted only those results found up to 1600, and no unambiguous usage was identified before 1550, thus searching up to about 30,000 books (EEBO TCP's Phase I and part of Phase II, as of October 2013).⁴ These books are of varying lengths and the majority (for example, all of Shakespeare's works) did not contain usage of *try-and-V*. Therefore the results are not presented as being statistically representative, and are used primarily for illustrative purposes. Additional limitations for a statistical approach to these results are discussed below.
- 24 Tottie's (2012) results showed that the sequence *try and [verb]* was used with a frequency of about 1.3 instances per million words from 1500–1549, up to about 2 instances per million words around 1600, and back down to around 1.1 instances per million words by 1700. Conversely, Tottie found no instances of the sequence *try to [verb]* from the early 1500s, but a few instances by the end of the 1500s with gradually increasing frequency up to the late 1600s at which point it surpassed *try and [verb]*

(around 2 instances per million words), followed by a jump in frequency to around 8 instances per million words around 1700. Thus *try and*[*verb*] was significantly more common than *try to* [*verb*] during the 1500s, and up until after 1650. We should apparently conclude from these results that *try-and-V* predates *try-to-V*, and that *try-and-V* remained the dominant construction for over 150 years, but these data points are fundamentally misleading due to the ambiguity of ‘try and’ during that time period. Although Tottie (2012: 209, footnote 11) did manually remove obviously irrelevant results from the automated search of EEBO, it is clear that the results were not *semantically* filtered manually such that only legitimate instances of *try-and-V* were included as opposed to literal coordination of the form *try and* [*verb*]. In fact, Tottie herself cites no verified examples of the *try-and-V* construction before 1550, despite presenting the frequency statistics as such. If the tabular data and graph were correct, then we should expect to find instances of *try-and-V* even *before* 1500, given that the construction is apparently already established during the early 1500s. No such examples are attested before 1550.

- 25 In a similar study, Ross (2013) also consulted the EEBO data from 1500-1600, having identified this period as the emergence of both the *try-and-V* and *try-to-V* constructions, given that there are no attested examples of either before 1550 but clear, if still infrequent, attestations by 1600. As in Tottie’s study, sequences of *try and* [*verb*] and *try to* [*verb*] were automatically collected from EEBO. However, each of these results was manually sorted based on interpretation: as a legitimate instance of either *try-and-V* or *try-to-V*, as a coincidental collocation, or as semantically and contextually ambiguous such that classification was uncertain. Consider the following examples, with (9) and (10) representing the first verified instances of the respective constructions, (11) and (12) representing excluded, coincidental collocations from the same time period, and (13) and (14) representing ambiguous instances.

26 **Verified results:**

- (9) You maie (said I) trie and bring him in, and shewe him to her.
(Heliodorus 1569, in EEBO)
(10) bycause the Romanes durst not trye to stope theyr passage.
(Watson 1568, in EEBO)

27 **Excluded results:**

- (11) to trye and proue what one is: *Aliquem inspicere experimentis*. [Latin: ‘examine someone with experiments’]
(Huloet 1572, in EEBO)
(12) [it] is tryed to knowe whither it haue any drosse in it
(Calvin 1574, in EEBO)

28 **Ambiguous results:**

- (13) In that therefore God suffereth ye greatest part of men to despyse his Gospell, & seketh not to tame them, he doth it to trye and exercise his ministers. (Marlorat 1570, in EEBO)
(14) to put downe the houses of Idols trye, to se that the lawe and worde of god be taught among ye people
(Coverdale 1547, in EEBO)

- 29 Given the extensive number of ambiguous results, Ross (2013) presented the entire set of sentences as an appendix, providing no conclusive analysis except to identify that the first instances of both constructions were found within one year of each other

(1568-1569), as presented above in (9) and (10).⁵ The following tables summarize the results presented in the appendix to Ross (2013):

Table 1. Summary of results from Ross (2013) appendix, 1500-1550

Type	Confirmed	Ambiguous	Excluded	Total
<i>try and</i>	0	14	18	32
<i>try to</i>	0	2	0	2

Table 2. Summary of results from Ross (2013) appendix, 1550-1600

Type	Confirmed	Ambiguous	Excluded	Total
<i>try and</i>	6	172	69	247
<i>try to</i>	34	4	6	45

- 30 In terms of methodology, the *ambiguous* results are most striking. Indeed, as Tottie reported, there are many more results for *try and* than *try to* in the EEBO data; however, the vast majority of these results are ambiguous, and only about 2% are verified as legitimate instances of the *try-and-V* pseudocoordination construction. Thus there are actually almost six times as many verified results of *try-to-V* as there as for *try-and-V* in this data set. The ambiguous results were coded as such conservatively, including many examples that were probably not pseudocoordination but could not be confidently excluded. Therefore, a minority of the ambiguous results are likely to represent legitimate instances of *try-and-V*. Even though the intuitions of English speakers today are not entirely reliable for their classification and thus the ambiguity may be potentially exaggerated by conservative analysis, it is this very ambiguity that marks the development of the construction. While most of the ambiguous examples are likely not instances of *try-and-V* per se, they may have contributed to its development during this time period as speakers began to occasionally interpret such usage as control-verb complementation.
- 31 In summary, then, *try-and-V* and *try-to-V* are attested from the same time in the corpus, and the claim made by Hommerberg & Tottie (2007) and Tottie (2012) that *try-and-V* predates *try-to-V* is refuted. There is no conclusive evidence that *try-and-V* is older or was more frequent at first because the majority of its instances were ambiguous during this period. We can only conclude that ambiguous instances with *and* were more frequent than ambiguous instances with *to*. However, Hommerberg & Tottie's assertion that both constructions have nearly 500 year histories is supported. As discussed in Ross (2013), in the centuries before this time period the lexical meaning of the verb *try* had shifted from 'test, experiment' to 'attempt', first in combination with nouns, and then during the 1500s this new lexical meaning extended to usage with verbal objects as well, resulting in both the *try-and-V* and *try-to-V* constructions. The *try-and-V* construction developed due to ambiguity with normal coordination, especially due to the frequent collocation of *try and examine* and other similar expressions in religious texts as in (11) or (13). The *try-to-V* construction developed out of purposive usage, in which *try* was likely originally transitive with an object that was tested in order to do

something else, as in (12) or (14). Both constructions are attested only after 1550 but were apparently conventional expressions of the language by 1600. Their relative frequency is difficult to determine due to the ambiguity of many examples, but we can observe that *try to* is more often attested in verifiable examples whereas *try and* maintains more ambiguous usage throughout the period, and while *try to* is less frequent, it is the ambiguous usage of *try and* that is most frequent, perhaps explaining the development of this unusual type of infinitival complementation in Early Modern English.

- 32 Given the degree of uncertainty imposed by the number of ambiguous examples, it is difficult to present statistical results, especially for this initial period. Section 2.2 presents the trajectory of development for the construction.

2.2. Later development of *try-and-V*

- 33 Having looked at the origin of *try-and-V* in Section 2.1, we can now consider its development until present day usage, as well as its relative frequency to *try-to-V* over time. By combining data both as reported by previous studies and directly from several corpora, a general trajectory of these constructions begins to emerge and can be tentatively illustrated.
- 34 Because *try-and-V* is significantly more frequent in British English, as discussed in Section 3, it is best to consider both American and British dialects independently following their shared origin. It is also worth emphasizing that, before time periods when records of spoken English become available, the results are limited to usage frequencies in written English; although likely following a similar trajectory, spoken usage for these colloquial phrases may have been higher.
- 35 We can begin by adjusting the results presented by Tottie (2012: 210) based on the semantic analysis in Ross (2013) discussed above. In short, although many matching strings are found in the corpus, few for *try and [verb]* are actually grammaticalized instances of *try-and-V* pseudocoordination. And as shown above, both *try-and-V* and *try-to-V* emerged between 1550-1600, while Tottie (2012: 210) showed a relative increase for *try-to-V* by 1700. The outline for the continued trajectory can be observed based on data from Google Ngrams (Michel et al. 2011):

Figure 1. *Try+verb* frequency in Google Ngrams⁶

- 36 Regarding the early development, especially during the 1500s, data is sporadic in Google Ngrams and has not been manually filtered, but at the very least it does confirm that there has been a clear increase in usage for both constructions during the last 500 years. Of course Google Ngrams is not a small corpus at all, but in this case that is a disadvantage because the data cannot be easily manually filtered (also due to raw data for most examples being inaccessible in the web interface).
- 37 If we separate the data from Google Ngrams to consider only the data from British English, the relative frequency of *try-and-V* is higher and lower for *try-to-V*, and vice versa for American English. The development for both dialects is parallel in overall shape, while British English shows a relative inclination toward *try-and-V* starting in 1800s, as shown by the much steeper transition from 1800-1900 for *try and* in the British English graph, as well as the somewhat lower frequency of current usage for *try to* in British English today.
- 38 The rise and then slight fall of *try-and-V* from 1500-1700 observed by Tottie is also found in the data from Google Ngrams. That shift is small in the graphs here due to scale, but regardless is unlikely to reflect legitimate usage of *try-and-V* and *try-to-V* without manual filtering. Instead, this trend probably indicates the gradual decline of the lexical meaning of *try* as 'test, examine' as frequent especially in religious contexts such as 'try and examine oneself' as in (11). Therefore the usage of *try-and-V* was consistently rising from 1550-1800 but not at the level of *try-to-V*.
- 39 In contrast to the data from Google Ngrams, Maia (2012: 10) reported no overall change in frequency for *try-and-V* from the period between 1810-2009 in the Corpus of Historical American English (COHA: Davies 2010). But it turns out this is due to an under-interpretation of the data in the corpus. Maia clustered the data to produce four time periods to compare and in doing so obfuscated important patterns. By looking at the results from COHA directly we can amend that interpretation:

Figure 2. *Try-and-V* frequency (pmw) each decade in COHA

- 40 Indeed, on average the frequency distribution seems to hover around 1-1.5 per million words throughout this time period, but the variance between the 1800s and 1900s is different, and more importantly the extremes in the 1800s can be explained. Aside from the data point for the 1810s, the frequency of usage appears to be generally increasing throughout the 1800s, which would be consistent with the data from Google Ngrams as

well. And in fact the initial segment of the graph, marked in gray, is an anomaly in the data: whereas the other sections of the corpus have 3 million or more words (over 10 million from the 1870s onwards), the section for the 1810s has only 0.64 million words. That data point for *try-and-V* is based on just a single token (compared to 9 tokens by the 1830s for example). This single token should not statistically skew the distribution to the extent that it does.⁷ As for the peak at the end of the 1800s, a similar but much less extreme rise and fall is also found in the Google Ngrams data. It is possible that this fall represents a response to the prescriptive attitudes disfavoring this construction, especially in American English, that came about during the late 1800s (cf. Ross 2013).

- 41 At the same time, the *try-to-V* construction also increases in frequency in the COHA data, more or less steadily rising from around 20 instances per million words to almost 120 by 1950, and flattening out after that.
- 42 As for British English, Uusi-Mäkelä (2013) confirms a general trajectory for an increasing usage of *try* with a meaning related to ‘effort’ (e.g., ‘attempt to’) for the period from 1710 to present day, with the *to*-infinitive complementation pattern being less frequent than an NP complement at the beginning of that period, but much more frequent by the early 1900s. And as the ‘effort’ meaning increased over the past 300 years, the other lexical meanings of *try* decreased substantially. In terms of the relative frequency of *try-and-V* and *try-to-V* Uusi-Mäkelä’s results diverge from the Google Ngrams data, specifically by finding 0 instances of *try-and-V* between 1710–1780, but finding a much narrower gap than above between *try-and-V* and *try-to-V* by 1900 until the present day. One reason for these discrepancies is probably the selection of texts included in the *Corpus of Late Modern English Texts* (CLMET: De Smet 2005) and Uusi-Mäkelä’s decision to represent the BNC with only the *Imaginative prose* sub-section, versus the more conservative, general purpose texts from EEBO and the books represented by Google Ngrams, as well as Uusi-Mäkelä’s relatively small sample size for *try+verb* tokens overall. Regardless, allowing for genre and register differences, the overall development is not incompatible with the pattern already described. Thus these differences emphasize the need to take register and genre into account as well as having a large sample of relevant results.
- 43 Based on a more complete search of the materials in the current, extended version of the corpus with a substantially larger data set (CLMET 3.0, resulting from a project by Diller, De Smet & Tyrkkö 2010), we see a clear trend of increased usage from the period between 1710–1920 for these constructions,⁸ as shown in the table below:

Table 3. Frequencies of uninflected *try+verb* in CLMET 3.0⁹

	1710–1780	1780–1850	1850–1920
try and	1 (0.1 pmw)	72 (6.4 pmw)	196 (15.5 pmw)
try to	181 (17.3 pmw)	332 (29.4 pmw)	911 (72.2 pmw)
<i>section size</i>	10,480,431 words	11,285,587 words	12,620,207 words

- 44 From a methodological perspective, as pointed out by an anonymous reviewer, we might want to consider frequencies (of relevant data) to be a better indicator than word count for measuring corpus size. Along these lines, it is worth noting that the vast

majority of the tokens of *try and [verb]* and *try to [verb]* from this time period are genuine examples of *try-and-V* and *try-to-V*, in contrast to the results from EEBO as discussed above. This makes sense because the overall frequency of *try* (and therefore those sequences) has increased well above the coincidental sequences found relatively often in the 1500s as the constructions had just entered the grammar. Therefore, while completely unreliable for the earliest usage of the constructions, the raw, not manually filtered results for *try and [verb]* and *try to [verb]* give representative frequencies for the construction today, which is convenient for using the larger corpora for this later time period, such as Google Ngrams.

- 45 That the usage of *try-to-V* continues to increase today is also shown by Collin (2014) based on data from the British newspaper *The Independent* from 1992-2009, with a rise of over 20 instances per million words from about 80 to over 100, though a steeper change than represented by the Google Ngrams data; likewise, *try-and-V* was also reported to be increasing, with a change of about 10 instances per million words, versus relative stability in the Google Ngrams data. This might thus indicate a register or genre shift in the newspaper more than a grammatical change in the language.¹⁰

Figure 3. Development of *try+verb (pmw)* in American English

Figure 4. Development of *try+verb* (pmw) in British English

- 46 These graphs are composite illustrations based on several sources. The origin is based on the data from EEBO, showing the first usage between 1550-1600. The outline for the continued trajectory is based on data from Google Ngrams, normalized to instances per million words based on current usage patterns according to COCA for written American English, and the British National Corpus (BNC) for written British English. The graphs have also been further adjusted to fit the most salient contributions of the other studies and data from additional corpora discussed above. Given that these graphs represent combinations of data sets, the *per million word* figures should be interpreted only as a general reference point, with emphasis on the overall development as a whole.
- 47 An unusual opportunity to look at the usage of *try-and-V* in spoken English is given by the Hansard corpus, which collects transcripts of 7.6 million speeches from the British Parliament from 1803-2005. Overall the development follows the path for written British English shown above, although the frequency of usage is higher overall:

Figure 5. Development of *try+verb* (pmw) in Hansard corpus

- 48 The differences represented in this data may be due to the formal usage of members of Parliament. The frequency does increase throughout the 1800s, but *try-and-V* lags behind a little. Both constructions see a substantial rise in frequency during the period from 1910-1950, which may be explained by the political atmosphere at the time due to the World Wars, a period marked both by a higher usage of *try-and-V* but also a wider range of complement verbs. For example, *try and overcome* appears especially often during this period. The frequency of both constructions stops increasing after 1950, while the decline during that period for *try-and-V* might be explained by the spread of prescriptive attitudes about the unacceptability of *try-and-V* described for American English above, or its informality of *try-and-V* in formal usage, which would be appropriate for encouraging wartime discourse but not formal usage.

2.3. Development of the modern *BFC*

- 49 The initial stage of development for the Bare Form Condition (BFC) is indicated by the earliest usage of *try-and-V*, outlined in Section 2.1. Among the six verifiable examples of *try-and-V* in EEBO (cf. the appendix to Ross 2013), four are infinitives (*to trie and proue your manhoode*; *to trie and learn*; *haue him try, and see...*; and *to trie and make the better sale*), another is an imperative (*Trie and see, quhither 3ee haue faith...*), and the other follows the possibility modal verb *may* (example 9 above). In the first place, none of this usage involves inflected verbs, but more importantly, these specific constructions afford a level of uncertainty fitting the ambiguity of the construction as it developed. None of these grammatical contexts pragmatically assert the second verb as factual. That is, infinitives, imperatives and possibility modals typically refer only indirectly to the attempted action, even if in the initial stages of grammaticalization the second verb would otherwise be grammatically entailed via literal coordination. Consider the following example:

(15) I hope you try and study syntax.

- 50 It is true that *I hope you try*, and also true that *I hope you study syntax*, regardless of whether this is interpreted as *try-and-V* pseudocoordination or as normal coordination. In the case of the *try-and-V* construction, only the first verb is entailed, because the second verb is embedded as a complement to the meaning of 'attempt to'. But in the case of literal coordination, the verb *hope* has already weakened the larger context so that neither *try* nor *study* is asserted directly. Therefore it does not matter whether this is interpreted as pseudocoordination or normal coordination because the most natural reading of either will have the same pragmatic effect.
- 51 Importantly, that ambiguity is not only what allowed for the grammaticalization of the construction, but also what marks the first stage of the BFC. This could not be observed without close analysis of these few verifiable examples of *try-and-V* among the many possible results in the corpus.
- 52 The next major development for the BFC came about with the increased usage of *try-and-V* in the 1800s. As discussed in more detail in Ross (2013), the first bare present-tense usage (cf. 1 above) was found in the 1800s. By this time usage in pragmatically assertive contexts had appeared, such as the type 'I will try and win' which necessarily would assert the second verb (> 'I will win') if normal coordination, and must therefore be unambiguously analyzed as complementation, not literal coordination. Earlier, *try-and-V* could be seen as merely an idiomatic paraphrase of *do by trying*, but with that lack

of assertion of the second verb, and especially the introduction of present-tense usage, by the 1800s the construction was instead a grammaticalized paraphrase of *try to do*:

(16) I shall try and get a good Husband.
(Chaigneau 1753: Google Ngrams)

- 53 With this cementing of the construction as distinct from its origin in normal coordination three centuries earlier, its Bare Form Condition was free to shift as well.¹¹ In the original usage, the BFC was merely a requirement that the construction only be used in non-finite (e.g., infinitival or imperatival¹²) contexts. But with reanalysis permitting present-tense usage, the BFC took on other properties: the requirement for a lack of inflection, but not necessarily non-finite forms:

(17) Perhaps, if you try and get well, I may comply with the wish which you so foolishly keep
(Burton 1838: COHA)

- 54 The final stage of development for the modern BFC is extending present-tense usage to permit *be* as a complement:¹³

(18) I try and be patient with her afore me, but I'd give aught I have to see her, and see what she wants.
(Gaskell 1848: CLMET)

- 55 Thus we have observed the *try-and-V* construction and its corresponding Bare Form Condition develop incrementally over almost 500 years.¹⁴

- 56 From a methodological perspective, it is important to observe that although large corpora are indeed available for English, including those consulted for this part of the current research, they are only barely sufficient for this level of detailed analysis for low-frequency phenomena. In fact, it seems to be the case that substantially more data is required for accurate observations of diachronic development when a construction originates out of ambiguity, as applies for most cases of syntactic reanalysis. Among a total of 279 examples of *try and [verb]* identified in EEBO between 1500-1600, only 6 were verified as legitimate instances of *try-and-V*, with various other results excluded as coincidental collocations but more importantly the vast majority of results found to be ambiguous and therefore unclassifiable. In other words, the explicated analysis above relied on a handful of examples among hundreds, for an already low-frequency phenomenon. Different historical corpora also introduce problems of register and genre, making measurements of overall frequency during the full 500-year period challenging. And the limitation to only written material, for which *try-and-V* is known to be significantly less frequent, may delay observable results from earlier changes in spoken usage. This portion of the research can still be considered a success because enough critical examples were identified to mark major developments, and the diachronic stages have been elucidated, even though some details remain opaque. In this case, the obstacle is not in finding a large number of examples, but in finding contextually unambiguous examples in representative works that mark milestones in the development. Thus a larger corpus would not, in itself, provide a better answer, except to the degree that it would provide more opportunities for identifying critical examples of this low-frequency phenomenon.

3. Dialectal variation

- 57 In Section 2, a 500 year history for *try-and-V* was established, along with the long-term parallel development of the construction in British and American English, with the only significant difference being that *try-and-V* is now relatively more frequent in British English than American English, a trend that can be clearly observed since the 1800s. The BFC was established in a nascent form in the 1500s and reached its current grammatical status by the beginning of the 1900s. In this section, given this long-term development, the question is asked whether there exist systematic exceptions to the BFC in any dialects of English.
- 58 Previous corpus-based research on the *try-and-V* construction has primarily discussed dialectal differences between British and American English, as well as the usage of the construction in spoken vs. written registers, and in different grammatical contexts, most notably the Bare Form Condition (BFC).¹⁵
- 59 That *try-and-V* is significantly more common was mentioned by Biber et al. (1999: 739), but the first substantial corpus study to compare the two dialects was Hommerberg & Tottie (2007),¹⁶ discussed also in Tottie (2009: 343–349):

Table 4. *Try-and-V* vs. *try-to-V* (Hommerberg & Tottie 2007:48)

	British Spoken	British Written	American Spoken	American Written
try and	71%	24%	24%	5%

- 60 Another study by Maia (2012) similarly reported a higher frequency of *try-and-V* in British English.
- 61 As for other varieties of English, Brook & Tagliamonte (2016) present a detailed corpus study of Canadian and British English, with results roughly analogous to American and British English, and they add a sociolinguistic perspective by looking at vernacular community-based dialect data.
- 62 From a broader perspective, the massive GloWbE corpus (Davies 2013) allows for the comparison of in 20 English dialects around the world, using informal, online writing, and the frequencies of *try-and-V* and *try-to-V* can be compared, as presented in Ross (2013: 127).¹⁷

Figure 6. Percentage try-and-V vs. try-to-V in GloWbE¹⁸

- 63 The British (GB) dialect stands out as having the most frequent usage of *try-and-V* (29%), with closely the associated dialects of Australia, New Zealand and Ireland at nearly the same level. Intermediate dialects such as Canadian (17%) show some influence of American (US) English (14%). All dialects show usage of the construction, although those that use it less frequently than American English may be outliers for which the construction is not a local conventional feature, as in Bangladesh (BD) with the lowest usage (10%).
- 64 The massive dataset provided by GloWbE also offers insight into just how robust the BFC is across all 20 dialects. Exceptions to the BFC are 1000 times rarer than tokens that follow the restriction, while *try-to-V* tokens are mixed:

Table 5. Inflection of *try+verb* constructions in GloWbE

	<i>try-and-V</i>	<i>try-to-V</i>
Bare	67888 (7%)	282359 (30%)
Inflected	64 (.007%)	595195 (63%)

- 65 Of the 64 tokens violating the BFC, 46 had a bare second verb (e.g., *He tries and go*), possibly by analogy to *try-to-V*, while the other 20 had parallel inflection on both verbs (e.g., *He tries and goes*).¹⁹ It is possible that this apparent uncertainty about how to extend *try-and-V* to inflected contexts has contributed to the resilience of the BFC in general, with both constraints (parallelism and the second verb as a bare infinitive) contradicting each other in the possible inflected forms. Likewise, the other rare exceptions to the BFC mentioned in this paper are not categorical in this way either. There are no clear dialectal distributional patterns among the 64 tokens, and many or most are likely typos or other mistakes because the corpus represents informal usage online.²⁰
- 66 Returning to the theme of this volume, it is worth emphasizing that for nuanced questions like this, even a massive corpus with 1.9 billion words has generated only a tiny corpus of 64 tokens for consideration of this important question: GloWbE is in

effect a *small* corpus, relative to this question. And while the 64 results confirm that the BFC appears to be ubiquitous across English dialects, any further splitting of the dataset to look at variation, with so few tokens for each dialect, would not produce statistically significant results. So we must turn toward objectively smaller corpora to find a better way to address the question of variation.

3.1. Smaller regional and divergent dialects

- 67 Of course the large-scale approach of GloWbE is advantageous for incorporating a massive amount of data but probably biases that data toward standard English usage, even within the many international varieties represented, especially because it represents communication on the internet rather than verified local native speaker usage. With no regional exceptions or variation for the BFC identified in the GloWbE data, we might wonder whether there is such variation within smaller regional dialects. Additionally, several native speakers have mentioned to me, during the course of this research, that in their personal usage inflected *try-and-V* is either acceptable or marginal, not ungrammatical, but I have found no geographic patterns from this anecdotal evidence, so continued research on possible variation in the BFC beyond standard usage is desirable.
- 68 Non-standard dialects present their own challenges such as mixing with standard forms in actual usage so that a pure corpus (or natural acceptability judgments) are almost impossible to obtain. And conversely it is possible that *try-and-V* is seen as a colloquialism of standard English and therefore less likely to appear within regionally marked, non-standard speech much in the same way that it is less likely to appear in formal registers of standard English. Still, this is a relevant avenue for research on variation and exceptions to the BFC.
- 69 Several published descriptions of regional dialects document relevant variation in usage of pseudocoordination, while other dialects feature morphosyntactic contexts that could affect the realization of the BFC. These dialects are discussed in turn below, sometimes as desiderata for the development of future corpora, as some of these dialects are in general not represented by corpora of sufficient size for observing variation in (or sometimes any usage of) *try-and-V* at this time.
- 70 *South African English*: as discussed in Ross (2013: 125), examples can be found on the internet of South African English violating the BFC. However, such usage was not verified in GloWbE, so a narrower corpus, possibly of spoken usage, might later verify this observation or provide more evidence against this usage being found more often than in other dialects.
- 71 *Channel Islands English*: *go-and-V*, a different pseudocoordination construction, has shifted to allow non-parallel inflection for speakers of Jersey English (cf. Rosen 2014), as in *went and see* ('went and saw'). However, no corresponding instances of the type *tried and see* are found in the small, fieldwork-based corpus used for that research (Rosen, p.c.); a larger corpus or additional fieldwork would be required to determine whether non-parallel usage has extended from *go-and-V* to *try-and-V*.
- 72 *Northern Subject Rule*: found in dialects from northern England, Scotland, and Ireland, to even Newfoundland and Appalachia, the Northern Subject Rule describes the use of *-s* in the present-tense beyond just third-person singular (with some restrictions, cf. Klemola 2000, de Haas 2011, Clarke 1997). In dialects where *-s* has become a general

marker of the present-, would *try-and-V* be absent, or would it be permitted with inflection (as suggested by Ross 2013: 124)? Corpora documenting this vernacular feature with sufficient tokens of *try-and-V* for analysis have not been identified.

- 73 *Dorset English*: If a dialect were to provide additional morphology beyond that which is found in standard English, it could be used to better understand the nature of the BFC. Such an opportunity might be given by southwestern dialects in England which add an unusual suffix -y to intransitive verbs, while in transitive usage they are bare as in standard English (Gachelin 1991; Trudgill 2011: 73–74, 105–106; Barnes 1886: 25), as in (19). Examples of this are found in the writings of William Barnes who wrote in Dorset dialect; unfortunately no instances have been found in combination with *try-and-V*, although Barnes did use *try-and-V* as shown in (20). If indeed these could be combined, finding an example of such usage is unlikely because the dialectal form has all but died out.

(19) Can ye zew-y?

Can you sew-INTR

'Can you sew?'

(Barnes 1886: 25)

(20) Aye, if a young chap, woonce, had any wit

To try an' scrape together zome vew pound,

To buy some cows an' teäke a bit o' ground...

(Barnes 1879: 104)

- 74 *East Anglian English*: A final variant of the BFC is its extension to additional grammatical contexts in dialects with additional functions for uninflected verbs. An example is provided by Faarlund & Trudgill (1999) who report that for the British dialects of East Anglia where bare verbs may be used for third-person singular present-tense in general, such usage is also permitted for the *try-and-V* construction; their claim can be verified by dialectal corpus data, as in (21):

(21) ...so he try and put a bit more... you know...

(June, female adult from East Anglia, 1991: BNC spoken section)

- 75 Of course the BFC itself remains unchanged in such usage.²¹

- 76 *Scots*: having developed in parallel to Old and Middle English, and now mixed with Scottish English, Scots is another language in which *try-and-V* and possibly the BFC might be found. Although *try-and-V* is found and the BFC appears to be maintained, using corpora to understand its usage and properties is difficult. There are ambiguous examples parallel to those found for English during the 1500s, as in the following example from the Helsinki Corpus of Older Scots (1450–1700) (HCOS: Meurman-Solin 1995) as in (22), whereas clearer usage may be due to the influence of English, as in (23), from *A Dictionary of the Older Scottish Tongue* (DOST, *and*), but also cited in the more recent revisions to the *Oxford English Dictionary* (likewise in the entry for *and*):

(22) Let euery man, and let euery woman trye and examine them selues.

(Bruce 1590: HCOS)

(23) [The council] ordanis the thesaure to trye and speik with Jhonn Kyle.

(Records of the Burgh of Edinburgh 1599: DOST: Craigie 1937: 76)

- 77 The SCOTS corpus (Douglas 2003; Anderson & Corbett 2008; 4.5 million words) shows usage patterns in Modern Scots similar to Modern English. The BFC is maintained in both the written (77%) and spoken (23%) sections of the corpus; as in English, *try-and-V* is much more frequent in speech:

Table 6. Inflection of *try+verb* constructions in SCOTS

Written	<i>try-and-V</i>	<i>try-to-V</i>
Bare	56 (16 pmw)	278 (80 pmw)
Inflected	0	608 (175 pmw)
Spoken	<i>try-and-V</i>	<i>try-to-V</i>
Bare	196 (189 pmw)	56 (54 pmw)
Inflected	0	285 (275 pmw)

- 78 Whether *try-and-V* is a legitimate construction of Scots, however, is harder to determine. The SCOTS corpus is intentionally not divided into Scots vs. Scottish English because these varieties exist on a continuum (Douglas 2003: 26–27; Anderson & Corbett 2008: 21), so the data for *try-and-V* may only represent a feature of Scottish English,²² although some examples appear authentic:

(24) I hae tae try an change mi tongue so fock understands me.
 'I have to try and change my speech so folks understand me.'
 (Man from Shetland: SCOTS)

- 79 To unravel these complexities, a dedicated study taking discourse factors into account, as well as likely working directly with native speakers, would be required.
- 80 Finally, a corpus of extensive spontaneous speech by the same individuals even within otherwise standard dialect areas might reveal systematic exceptions to the BFC, if it turns out that the variation reported to me anecdotally by several native speakers is not associated with geographic dialects but possibly with individual variation. Along these lines, Brook & Tagliamonte (2016: 309–310) report that over 60% of the almost 500 speakers in their sociolinguistic interview corpora (representing Canadian and British English) were *categorical* in their selection of either the *try-and-V* or *try-to-V* variant (in contexts where *try-and-V* was permitted by the BFC). Thus frequencies reported for dialects may apply only on average, rather than representing individual usage patterns.²³ If similar individual variation were identified for the BFC, that would be of great importance to understanding the nature of this unusual morphosyntactic constraint, although based on how infrequent BFC violations are in corpora such speakers might be rare.²⁴
- 81 In summary, although *try-and-V* appears to be widespread and the BFC robust throughout at least the standard varieties of English around the world, variation in the BFC might still be observed given large enough corpora of variant dialects, especially those less associated with standard English, and also if there is individual variation.

4. Acquisition

- 82 Having established the Bare Form Condition (BFC) as the norm throughout the English-speaking world, we might wonder how easily and early it is learned during language acquisition. Given the BFC has been identified as a particularly unusual and complex component of English grammar (Ross 2014), along with the relatively unusual use of *and* in the construction, we might expect it to be difficult and marked by obvious signs of trial and error during acquisition, or that it would be acquired late. This section thus

investigates the acquisition of the BFC by children acquiring English as their native language, and Section 4.1 is a brief discussion of second-language acquisition.

- 83 To understand the development of the BFC during acquisition and test for statistically significant distributions, a single corpus with sufficient tokens of *try-and-V* is required. Many corpora with data from children are limited in size for practical reasons. A useful resource for such research topics is the CHILDES database (MacWhinney 2000), which contains dozens of corpora for English and other languages. However, most of the corpora are quite small and contain no more than two instances of *try-and-V*. Because the input frequency for *try-and-V* is significantly higher in spoken British English, the corpora consisting of data from British children will be more likely to contain a sufficient number of tokens. Only two such corpora were found to be large enough for this study.
- 84 The first of these was the Fletcher corpus (Fletcher & Garman 1988; Johnson 1986) with cross-sectional data from 72 children ages 3, 5 and 7. As shown in Table 7, the *try-and-V* construction was produced by some children aged 5 and 7:

Table 7. *try+verb* constructions in the Fletcher corpus

(By child, with total instances in parentheses)

3 years	<i>try-and-V</i>	<i>try-to-V</i>
Bare	0	0
Inflected	0	4 (6)
5 years	<i>try-and-V</i>	<i>try-to-V</i>
Bare	2	0
Inflected	0	6 (10)
7 years	<i>try-and-V</i>	<i>try-to-V</i>
Bare	4 (8)	0
Inflected	0	6 (12)

- 85 *Try-and-V* was not produced by 3 year old children in this corpus, but with so few tokens overall that is not conclusive. However, these results are sufficient to support a statistical test: Fisher's exact test revealed a statistically significant correlation between presence of inflection vs. construction type ($p < .05$ for 5-7 years; there are too few tokens at 3 years to meaningfully apply the statistical test).²⁵ This test of course does not indicate that inflected *try-and-V* is ungrammatical or would never be produced, but that the correlation shown in the table is unlikely to be due to chance alone, supporting the interpretation of a grammatical difference between the constructions.
- 86 To complement those cross-sectional results, the Thomas corpus (Lieven, Salomo & Tomasello 2009) provides dense longitudinal data for a single child, recorded weekly at age 2, then monthly at ages 3 and 4. The results show that Thomas acquired the BFC early and consistently:

Table 8. *try+verb* constructions in the Thomas corpus

2 years	<i>try-and-V</i>	<i>try-to-V</i>
Bare	2	0
Inflected	0	3
3 years	<i>try-and-V</i>	<i>try-to-V</i>
Bare	6	5
Inflected	0	35
4 years	<i>try-and-V</i>	<i>try-to-V</i>
Bare	15	3
Inflected	0	31

- 87 There are no violations of the BFC, and the lack of inflected *try-and-V* for ages 3 and 4 is statistically significant ($p < .001$ with Fisher's exact test; age 2 is not quite significant at $p < .1$). Most surprisingly, however, may be that Thomas begins to use *try-and-V* before 36 months. The unusual grammatical nature of the BFC does not seem to delay acquisition of the construction.
- 88 Remarkably, Thomas does in fact make two errors with *try+verb* yet does not violate the BFC. The first lacks subject-verb agreement and therefore allows *try-and-V*, while the second is inflected but found with *to* rather than *and-V*, conforming precisely to the BFC:
- (25) It feels like a crab try and get you.
(Thomas: 4 years, 7 mo.)
- (26) I trying to concentrate.
(Thomas: 2 years, 11 mo.)
- 89 The results from these two corpora also suggest a categorical difference in early acquisition: *try-and-V* is exclusively uninflected, while *try-to-V* is almost exclusively inflected (similar some speakers of British English reported by Brook & Tagliamonte 2016 as discussed in Section 3.1). This may hint at how the construction is acquired: *to* when inflection is present, but *and* when it is not. The only exceptions are in Thomas's speech at 3-4 years old, presumably after having already established these constructions categorically and then adjusting their grammatical properties slightly, arriving at adult-like levels by age 4. In a way, this mirrors their diachronic development as well, with each construction at first used in a narrow domain then extended to more general usage. Note that for adults *try-and-V* is the more common form in spoken British English when both verbs are uninflected, but *try-to-V* still accounts for a substantial minority of the usage (for example Hommerberg & Tottie 2007: 48 report 29% for *to* and 71% for *and*). In fact, the children seem to acquire *try-and-V* as the default while *try-to-V* is used when *try-and-V* cannot be used because there is an inflected verb. We can also infer from this situation that the construction is acquired early in children hearing British English as input because the construction is so frequent in spoken British English, and they are not yet exposed to prescriptive or formal biases against *try-and-V* in writing.

- 90 While a similar effect of the construction being more common in spoken (vs. written) American English, thus contributing in both dialects to its ongoing preservation from one generation to the next, the relatively low frequency in spoken American English (Hommerberg & Tottie 2007: 48 report 29% for *to* and 71% for *and*) might suggest that American children would learn the construction later or less categorically. Regardless of these potentially different paths of acquisition, the BFC emerges for children in both dialects.
- 91 Turning now to the best candidate among the limited options for finding multiple tokens of *try-and-V* among the American English corpora available in CHILDES, the Gillam corpus (Gillam & Pearson 2004) offers a relatively high number of tokens for *try-and-V* to analyze. Short narrative stories were collected from 520 children ages 5-11.²⁶

Table 9. Frequencies of *try+verb* in Gillam corpus by year²⁷

(By child, with total number of instances in parentheses)

	5	6	7	8	9	10	11
try and	1	0	1	0	1	2 (3)	0
tries and	0	0	0	0	1	0	0
tried and	0	0	0	0	0	0	0
trying and	0	0	0	0	0	0	0
try to	1	1	2	4 ²⁸	2 (3)	2 (3)	2
tries to	0	0	1	1	4 (7)	1	3
tried to ²⁹	10 (16)	22 (32)	34 (46)	31 (50)	22 (33)	14 (25)	22 (27)
trying to	8 (9)	21 (24)	20 (22)	23 (29)	7 (10)	11 (17)	15 (18)

- 92 This data provides a few tokens of *try-and-V* along with many for *try-to-V*, reflecting that *try-and-V* is the much less frequent construction in American English. Unfortunately older children are represented in this corpus than the corresponding British children discussed above, and the data is not dense enough to identify developmental any changes, with the instances of *try-and-V* dispersed from 5 to 10 years of age. Setting aside age, the relationship between inflection and construction type appears to be statistically significant ($p < .001$ with Fisher's exact test) but this result is skewed substantially by the high number of past tense tokens in the stories told by the children, which is a factor not due to chance but stimulus design for collecting the corpus material. Therefore, this statistically significant result must be interpreted very cautiously.³⁰

Table 10. *try+verb* constructions in the Gillam corpus

(By child, with total number of instances in parentheses)

	<i>try-and-V</i>	<i>try-to-V</i>
Bare	5 (6)	14 (16)

Inflected	1	270 (375)
-----------	---	-----------

- 93 It can also be noted that in contrast to the tokens produced by the British children, there is no tendency toward a categorical difference between *try-and-V* as exclusively bare and *try-to-V* as exclusively inflected. This reflects the lower frequency of *try-and-V* in adult American speech.
- 94 There is another important difference from the data for the British children: one token appears to violate the BFC:
- (27) The girl wants to go over there to meet them. But the boy says no. When the girl *tries and tugs* him over there the boy says no again and tries to stop her.
(Male child in Gillam corpus: 9 years, 4 months)
- 95 Although this sentence is ambiguous,³¹ the ‘attempt to’ interpretation seems most plausible in context. Compared to only 6 instances of uninflected *try-and-V* in the corpus, this single likely exception to the BFC is not insignificant. The American English data therefore may be more relevant to studying variation and exceptions to the BFC, specifically because *try-and-V* is less frequent and therefore fewer tokens are provided to children as input. Despite having less easily accessible data with fewer instances of *try-and-V*, an acquisition corpus for American English may actually be more informative for the research questions posed here. Thus an effectively smaller corpus is a better corpus, in a way, although it would still be desirable to have more data.
- 96 More broadly, the acquisition of the BFC can be considered within the context of the Grammatical Conservativity Hypothesis (Snyder 2007, 2008, 2011), which states that children will often make errors of omission, but few of *comission* (producing elements not found in adults speech). Grammatical conservativity suggests that children will not overgeneralize and produce inflected *try and*, while the poverty of the stimulus argument demonstrates that children must generalize from impoverished input: if children do not produce inflected forms, this must be because they acquire the BFC, a conclusion supported by the corpus results. These results are parallel to those found in another study about Grammatical Conservativity for the acquisition of the *go get* (or *come see*) construction, which has an at least superficially similar bare form requirement.³² Sugisaki & Snyder (2013) found in a corpus of speech from children ages 2-5 only 3 inflected instances of the construction, versus 439 uninflected instances. This shows a strong preference among young children to produce only utterances that do not violate grammatical constraints, thus supporting the Grammatical Conservativity Hypothesis. These results are very similar to the results presented above for *try-and-V*. Even young children have internalized the restricted inflection of these constructions.
- 97 The conclusion we must draw from this is that whatever underlying grammatical properties explain the BFC, they are acquired early and consistently by children.
- 98 Another question is whether the *try-and-V* construction is learned primarily as a productive grammatical device or as an idiomatic construction associated with a small set of complement verbs that is generalized by adulthood. In order to answer this question, we can look at the lexical collocations of the construction. Gries & Stefanowitsch (2004: 123) present the following lists of verbs that are most attracted to the *try-and-V* versus *try-to-V* constructions from a collostructional analysis:³³

Table 11. Collexemes by frequency (Gries & Stefanowitsch 2004)³⁴

try and	get , come, teach, show, learn, set, see, bring, be
try to	make , do, put, analyse, give, improve, persuade, use, keep

- 99 Are these also the most frequent lexical collocations for *try-and-V* during acquisition? Surprisingly, that does not appear to be the case for the corpora consulted above:

Table 12. Lexical tokens from *try-and-V* in child corpora.

Thomas (age 2-4)	close, roll; get, hide, make, put (2), shout; catch (2), do, eat (3), fit, get (3), open, put, sell, work, (visit?)
Fletcher (ages 5-7)	duck, jump; catch, do, finish, fit, get, knock, tick, bunny-hop
Gillam (ages 5-11)	catch (2), find, fix, run, see

- 100 Although several verbs do appear in both lists, such as the most frequent verb *get* (5 tokens in the child corpora), the lexical profile in these tables is actually quite distinct. A full collocation analysis (as in Stefanowitsch & Gries 2003, Gries & Stefanowitsch 2004, or Wulff 2006) would require a much larger data set of *try-and-V* tokens to have statistically meaningful results, especially to observe any shifting patterns at different ages, but the data above suggests that children acquire *try-and-V* as a productive grammatical pattern at an early age. In fact, it is the older children (and Thomas by age 4) who appear to have the most adultlike usage, whereas the younger children appear to use *try-and-V* in creative, childlike expressions, such as the following example:

(28) ...if I go into the curb I try and bunny-hop up it.

(Male child in Fletcher corpus: 6 years, 11 months)

- 101 The two tokens that Thomas produced of *try-and-V* at age 2 were *roll* (2 years, 9 months, 9 days) and *close* (2 years, 9 months, 24 days),³⁵ which are not among the identified collocates for adult usage. This evidence strongly suggests that children acquire *try-and-V* as a productive grammatical device. It is possible that *try-and-V* is originally learned in idiomatic phrases such as ‘try and get’ or ‘try and see’, but the fact that the corpora lack early or unpredictably frequent usage of these verbs indicates that the children are quickly capable of generalizing the construction for productive usage in combination with any verb. Thus the uninflected form of the verbs cannot be idiomatically memorized either, given that the construction is productive, so at least some version of the BFC is acquired early and easily by children. Additionally, by adulthood they have further extended the construction to usage in a full range of grammatical contexts, as well as for use with other first verbs beyond *try* such as *be sure* and *remember*, but addressing that question would require more data because there are no tokens of *be-sure-and-V* or *remember-and-V* in any of these three corpora. However, we can look at the range of grammatical contexts found in the data from the child corpora:

Table 13. Grammatical contexts for *try-and-V* in child corpora³⁶

Thomas (age 2-4)	INF, IMP; TO (3), INF (2), CIT; IMP (5), TO (4), INF (2), CIT, PRES
Fletcher (ages 5-7)	INF, PRES; TO (3), PRES (2), INF, COORD [larger context: TO], CIT
Gillam (ages 5-11)	TO (4), INF, COORD [larger context: TO]

- 102 We can observe a wide variety of grammatical contexts for *try-and-V* in this data, not unlike the distribution for adults (cf. Lind 1983). Especially common are tokens in imperatives and infinitives (with or without *to*), and the earliest tokens produced by Thomas age 2 are an infinitive and an imperative, much like the diachronic origin of the construction (cf. Section 2). The principle of *horror aequi* (cf. Rohdenburg 2003, Biber et al. 1999: 738–739) could be a contributing factor but not a strong rule in the children’s grammar: 14 tokens are *to*-infinitives (*to try and...*), while only 2 are coordinated (*and try and...*). As for the BFC, 35 of the tokens are non-finite forms, versus only 4 tokens in the present tense; one is Thomas age 4, and the others are from one 5 year old and two 7 year olds in the Fletcher corpus. This could indicate that extension of the BFC to the present tense is a later development during acquisition, but the evidence is inconclusive. The single apparent violation of the BFC, mentioned above as (27) but excluded from the results in the table, might be used as anecdotal evidence that extending the construction to the present tense could pose a temporary problem for some children during reanalysis and possibly lead to overextension, but we cannot make any conclusions from this single token.

4.1. Second language acquisition of the BFC?

- 103 We now briefly turn to the question of second language acquisition. Do adults learning English as a second language learn the BFC as easily and quickly as children do? A full study of this research question would require extensive corpus data separated, including separate data sets for different language backgrounds. For example, given that in some Swedish and Norwegian dialects the equivalent ‘try and’ construction permits inflection (e.g., lit. ‘He tried and did it’ for ‘tried to do’, cf. Ross 2015), it would be interesting to determine whether transfer plays a substantial role during acquisition. And adult learners are more likely to be affected by prescriptive biases against *try-and-V*, and may even be explicitly warned against using the construction by usage guides, or simply not exposed to it as an alternative to *try-to-V*.
- 104 Below the results of an exploratory study based on the Michigan Corpus of Academic Spoken English (MiCASE) are presented to try and address this question in general. The corpus consists of transcribed spoken academic interactions on an American university campus, and the transcripts are divided by speaker background, permitting a comparison of American Native speakers, Near-Native speakers (who are fluent non-native speakers whose dominant language is English), and Non-Native speakers (excluding the Near-Native speakers).

Table 14. Frequencies of uninflected *try+verb* in MiCASE

	American Native	Near-Native	Non-Native
try and	98 (53 pmw)	2 (5 pmw)	0
try to	387 (209 pmw)	36 (93 pmw)	35 (68 pmw)
<i>section size</i>	1,847,433 words	387,101 words	514,923 words

Table 15. Frequencies of *try+verb* with inflection in MiCASE

	American Native	Near-Native	Non-Native
try+INFL and	0	0	0
try+INFL to	866 (469 pmw)	98 (253 pmw)	37 (72 pmw)
<i>section size</i>	1,847,433 words	387,101 words	514,923 words

- 105 Several observations are obvious from these tables. The first is that no tokens of inflected *try-and-V* were produced by any speakers, whether native or non-native. The second is that no instances of *try-and-V* were produced by the Non-Native speakers and only 2 (or 5 per million words) were produced by the Near-Native speakers, compared to 98 (53 per million words) by the American Native speakers. The third is that overall *try+verb* complementation is used much more frequently by American Native speakers than either Non-Native group, while the Near-Native speakers use *try+verb* more often than the lower proficiency Non-Native speakers. Thus apparently frequent use of *try* is a marker of native-like speech in this corpus.
- 106 But why do the Non-Native speakers not use *try-and-V*? The variant appears to be used about 20% of the time in American academic English according to the data from native speakers. Several contributing factors are likely: as a minority variant, *try-and-V* is simply not needed by non-native speakers; there may be a prescriptive bias against the construction or it may not be introduced in textbooks; the BFC (or the use of *and* in this construction) may be confusing, such that English learners avoid it. Furthermore, this corpus might represent different registers for native and non-native speakers because the native speakers may feel more comfortable appropriately using colloquial phrases that the non-native speakers would avoid in academic discourse. A combination of these factors is the likely explanation, and these results warrant further research. The current data set does not permit a finer-grained analysis, and more tokens, especially from a range of proficiency levels and in different registers, would be required to fully explicate these results. Any effects of the learner's first language are also unclear from these results with too few tokens for *try-and-V*. Thus, this low-frequency phenomenon is of even lower frequency in a learner corpus perhaps 100 instances of *try-and-V* in native speech for every one instance in non-native speech. In fact, as shown in the data from near-native English speakers, it is possible that some learners never learn to produce the construction, even after reaching high levels of proficiency, although they probably recognize it when listening. This is in sharp contrast to the patterns described above for first-language development, where despite its lower frequency, *try-and-V* is learned reliably alongside *try-to-V* and at first almost categorically for uninflected

try+verb contexts at least in British English. A corpus of British English second language speakers might also reveal dialectal differences in this regard as discussed above for first-language acquisition.

- 107 The only other corpus study to consider *try-and-V* usage in non-native speakers is Ruuska (2013), who investigated the relatively informal writing of online blogs written by Swedish first-language speakers for whom English was a second language in a corpus of 75,393 words, and by English speakers (presumably mostly native speakers) in the Birmingham Blog Corpus (created at but not limited to speakers from Birmingham University) with over 628 million words:

Table 16. *Try+verb* in blog corpus data (Ruuska 2013: 26–27)

	Birmingham Blog Corpus	Swedish L1 Blogs
try and	21125 (0.3 pmw)	1 (0.1 pmw)
try to	92646 (1.5 pmw)	25 (3.3 pmw)

- 108 Ruuska suggested that this low frequency of usage of *try-and-V* in both the Sweden and Birmingham blogs could indicate the influence of American English where *try-to-V* is more dominant than in British English. However, given that blogs are a written format, it is not surprising to find 19% usage of *try-and-V* in the Birmingham blogs, compared to the 24% baseline reported by Hommerberg & Tottie (2007: 48). The Birmingham Blog Corpus is over 8,000 times larger than the corpus of blogs written by Swedish first-language speakers, and with only a single example, we cannot make any generalizations about the acquisition of *try-and-V*. (It can be noted that this single example did not violate the BFC, however.) Still, these results appear to reinforce the results from the MiCASE corpus above, suggesting that non-native speakers rarely use *try-and-V* even at high proficiency levels. To strongly make this generalization, however, a large corpus of informal spoken non-native English should be consulted; that is, not informal writing as in the Sweden blogs corpus, nor academic (and therefore relatively formal) spoken English as in MiCASE.³⁷

Conclusion

- 109 *Try-and-V* pseudocoordination is truly a low-frequency phenomenon in English, and its investigation in corpus data requires selecting examples from subsets of subsets: those uses of the verb *try* followed by a verbal complement linked by *and*. And specifically to study the Bare Form Condition from various perspectives —diachronic, dialectal variation, and acquisition— this data is further reduced to even a subset of that data. Specifically, we must look beyond the most frequent instances of *try-and-V* as found in infinitives and imperatives, because they are inherently uninflected, and try and determine instead whether finite usage, as in the present tense, is extended to any inflected forms within the sample.
- 110 As shown above, extracting relevant results from corpora with millions —or even billions— of words can still be challenging, resulting in a different perspective on what a ‘small corpus’ is. The goal is of course simply to identify a corpus with sufficient data to address the research questions, but in this way low-frequency phenomena can make

even relatively large corpora seem small. Low-frequency phenomena such as *try-and-V* present unique challenges to corpus design and use, yet low-frequency phenomena are at least as important for grammatical theory as more frequent features (Ross 2014).

- 111 This is most acutely problematic for domains where available data sets are inherently limited for practical reasons, such as the acquisition data from a single child, or the colloquial speech of small dialects. Yet the outlook in this article is overall optimistic, given that the state of the art of current corpora have supported successful analysis of this low-frequency phenomenon and sufficiently addressed the main research questions, though not without challenges. Several more specific questions that emerged, or were not fully addressed, remain targets for future study, to be supported by larger corpora for narrower domains than are currently available. Still, this broad corpus research has clarified the profile of the *try-and-V* construction in English, as it developed, as it varies in major dialects today, and as it is acquired.
- 112 Research on a specific syntactic construction based on data from only a single, though frequent, verb is possible but difficult in English, but one important domain introduced at the beginning of this article remains to be addressed. Corpora for other languages lag behind, when compared to the massive databases that have recently emerged for English. And questions about diachronic development, dialectal variation and acquisition remain out of reach for most languages today. It is worth turning our attention to a parallel ‘try and’ construction in Faroese (North Germanic, approximately 65,000-70,000 speakers). The curious development of *try-and-V* in English appears to be emerging in Faroese as well (Heycock & Petersen 2012; Ross 2015). Along with *royna at [verb]* ‘try to’, Faroese has an alternative construction *royna og [verb]* ‘try and’. In normal usage this construction is limited to three non-finite morphological forms: the singular and plural imperatives, and the infinitive, as in (29):

(29)

Tú	mást	royn- a	og	les-a	bókina.
You	must- PRES. 2SG	try- INF	and	read- INF	book.DEF
‘You must try and read the book.’ (Faroese: Heycock & Petersen 2012: 275)					

- 113 The Faroese construction is therefore in essentially the same distribution as English from 1550-1800 (see Section 2). However, Heycock & Petersen (2012: 274-275) mention in a footnote that some speakers also accept this construction in the plural present tense. Just as English present-tense verbs, aside from third-person singular, are identical to the infinitive form, in Faroese the plural present tense forms are identical to the Faroese infinitive, ending in *-a*. (The singular forms display distinct subject-verb agreement morphology.) Therefore, extension to a subset of finite forms is possible if speakers accept them because they *look like* infinitives, as in the following example reported by Heycock & Petersen to be accepted by some speakers, parallel to the extension of English *try-and-V* to the bare present-tense that occurred in the 1800s:

(30)

%Tey	royn-a	og	les-a	bókina.
------	--------	----	-------	---------

They	try- PRES.PL	and	read- INF-PRES.PL	book.DEF
'They try and read the book.' (Faroese: Heycock & Petersen 2012: 274)				

- 114 Note that this Faroese equivalent to the 'Bare Form Condition' is not morphologically bare at all but instead requires that a verb look like a non-finite form. Such a development, given the richer and more informative morphology in Faroese would also by analogy support the infinitive-based analysis of the BFC proposed above and in Ross (2014), making this data valuable in allowing us to observe this diachronic change in progress.
- 115 Unfortunately several factors impede corpus research on Faroese 'try and': the construction appears to be less frequent than in English, and possibly more prescriptively stigmatized, so there would be fewer tokens in a corpus of comparable size, and the available corpora for Faroese are nowhere near as large as for English.³⁸
- 116 The most promising candidate is Føroyskt TekstaSavn comprised of about 4 million words from the 1998 year of the Faroese newspaper *Dimmenlætting* (Hansen 2003, 2005; Rasmak 2005). The results were unfortunately minimally informative, however. Only 10 tokens of *royna og [verb]* were identified, including 9 imperatives (6 singular and 3 plural) and one infinitive. There were no results with finite forms or other inflections. These results do strongly suggest that Faroese 'try and' is restricted as reported by Heycock & Petersen (2012). That is, finite usage is not permitted in general. However, the results do not provide evidence for an extension to present-tense plural forms that would mirror the modern English BFC. If this usage is only on the verge of developing and accepted by a minority of speakers, then finding no examples among 10 tokens should not be surprising and cannot be considered counter-evidence either.
- 117 Meanwhile, the corpus results do introduce another interesting difference between English and Faroese. 90% of the tokens identified in the corpus are imperatives, not infinitives, and even with the small sample size, this suggests a fundamental difference compared to English usage where infinitives are by far more frequent. Hommerberg & Tottie (2007: 50–51) report up between 4 to 12 times more infinitives than imperatives in English (varying by written vs. spoken registers and American vs. British English). Additional corpus data from Faroese to investigate the effects of register would be helpful, given that newspaper texts are relatively formal and therefore might support mostly colloquial imperative examples in quotations, for example.
- 118 We are left with uncertainty about Faroese 'try and' from the corpus data alone, especially regarding the inflectional restrictions, so the question must instead be addressed by grammaticality judgments from native speakers as in Ross (2015), where corroborative evidence was found in the form of intermediate acceptability ratings, between the levels of non-finite usage and (ungrammatical) finite inflections such as the past tense. The Faroese speakers also rated imperatives as significantly more acceptable than infinitives, whereas English speakers rate them at the same level, suggesting that the corpus results with 9 imperatives and only 1 infinitive may be representative rather than an effect of register.
- 119 The outlook of this article remains optimistic because as corpus technologies and available data continue to improve, low-frequency phenomena should become accessible for study in other languages as well, just as they are now accessible in the

best corpora available for English today. Holistic pictures of syntactic constructions, beyond contemporary, standard English, are within our reach. The results of this case study are encouraging and question the easily made assumption that bigger, general corpora are better: for addressing many important questions we need diverse corpora—with as much data as possible, but also representing as many languages and domains of usage as possible.

- 120 Central issues in corpus research methodology are coverage of diverse domains (as shown by the current case study), representativeness (see also Raccach, this volume, on the empirical requirements for scientific contributions in corpus research and otherwise), relevance to the study (for example, Debras, this volume, who required specialized corpus data aligning speech and gesture), and generalizability (see also Hennecke, this volume, on making individual corpus studies comparable). Small corpora are especially suitable for many of those purposes, and often have advantages over large corpora. The current study highlights one advantage of small corpora in particular, namely that manual analysis of the data is possible. Pseudocoordination presents a challenge for automated corpus research: the word *and* will likely never be automatically tagged correctly in instances of pseudocoordination, and the raw results are inherently ambiguous before being interpreted in context. As shown in Section 2, large, automatically generated data sets may produce misleading results before manual correction, especially during periods of change. As shown in Section 3, small, specialized corpora may better address questions of variation than large, general purpose corpora. And as shown in Section 4, even case studies of individual children can provide potential generalizations about grammatical structure. Still, as shown for Faroese above, and elsewhere in this case study, the lack of *sufficiently* large corpora can still be an limiting factor. Regarding corpus size and defining ‘small corpora’ in general, for the future of effective linguistics research, it is more important not whether we have large or small corpora in absolute terms but that we have a variety of appropriate corpora that are *large enough* to address a variety of important questions and make relevant contributions to our understanding of language use and the human language faculty.

BIBLIOGRAPHY

Corpora

BNC (2007). *The British National Corpus*, version 3 (BNC XML Edition). Distributed by Oxford University Computing Services on behalf of the BNC Consortium. <http://www.natcorp.ox.ac.uk/>

CHILDES. Child Language Data Exchange System. <http://childes.psy.cmu.edu/>

CLMET. Corpus of Late Modern English Texts. <https://perswww.kuleuven.be/~u0044428/clmet.htm>

CLMET 3.0. Corpus of Late Modern English Texts, version 3.0.

https://perswww.kuleuven.be/~u0044428/clmet3_0.htm

- Davies M. (2008). The Corpus of Contemporary American English (COCA): 450 million words, 1990-present. <http://corpus.byu.edu/coca/>
- Davies, M. (2010). The Corpus of Historical American English (COHA): 400 million words, 1810-2009. <http://corpus.byu.edu/coha/>
- Davies M. (2013). Corpus of Global Web-Based English (GloWbE): 1.9 billion words from speakers in 20 countries. <http://corpus2.byu.edu/glowbe/>
- EEBO TCP. Early English Books Online - Text Creation Partnership. <http://www.textcreationpartnership.org/tcp-eebo/>
- Føroyskt TekstaSavn. Faroese text collection by Sprákbanken and Fróðskaparsetur Føroya. <http://spraakbanken.gu.se/FTS/>
- Google Ngrams. (Michel et al. 2011). <http://books.google.com/ngrams/>
- HCOS. Helsinki Corpus of Older Scots. Department of Modern Languages, University of Helsinki. <http://www.helsinki.fi/varieng/CoRD/corpora/HCOS/index.html>
- MiCASE. (2007). *Michigan Corpus of Academic Spoken English*. Ann Arbor, Michigan: The Regents of the University of Michigan. <http://quod.lib.umich.edu/cgi/c/corpus/corpus>
- Nordic Dialect Corpus. <http://www.tekstlab.uio.no/scandiasyn/>
- SCOTS. Scottish Corpus of Texts & Speech. <http://www.scottishcorpus.ac.uk/>

References

- Anderson W., & Corbett J. B. (2008). "The Scottish corpus of texts and speech - a user's guide", *Scottish Language* 27, 19–41.
- Barnes W. (1879). *Poems of Rural Life in the Dorset dialect*. London: C. Kegan Paul & Co.
- Barnes W. (1886). *A Glossary of the Dorset Dialect with a Grammar of its Word Shapening and Wording*. Dorchester: M & E Case.
- Biber D., Johansson S., Leech G., Conrad S. & Finegan E. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.
- Bjorkman B. M. (2016). "Go get, come see: Motion verbs, morphological restrictions, and syncretism", *Natural Language & Linguistic Theory* 34(1), 53–91.
- Brook M. & Tagliamonte S. A. (2016). "Why Does Canadian English Use *try to* but British English Use *try and*? Let's Try and/to Figure It Out", *American Speech* 91(3), 301–326.
- Bryant M. M. (ed.). (1962). *Current American Usage*. New York: Funk & Wagnalls.
- Carden G. & Pesetsky D. (1977). "Double-Verb Constructions, Markedness, and a Fake Coordination", *Chicago Linguistics Society* 13, 82–92.
- Clarke S. (1997). "English verbal -S revisited: the evidence from Newfoundland", *American Speech* 72(3), 227–259.

- Collin C. (2014). "Hendiadys et construction de l'événement en anglais contemporain (Hendiadys and construction of an event in contemporary English)", *Neophilologica* 26, 90–105.
- Craigie W. A. (ed.). (1937). *A Dictionary of the Older Scottish Tongue: From the Twelfth Century to the end of the Seventeenth. Volume I: A-C*. Chicago & London: University of Chicago Press & Oxford University Press.
- Davies M. (2014). "Examining Syntactic Variation in English: The Importance of Corpus Design and Corpus Size", *English Language and Linguistics* 19(3), 1–39.
- Davies M. & Fuchs R. (2015). "Expanding horizons in the study of World Englishes with the 1.9 billion word Global Web-based English Corpus (GloWbE)", *English World-Wide* 36(1), 1–28.
- De Haas N. K. (2011). *Morphosyntactic variation in Northern English: the Northern subject rule, its origins and early history*. LOT, Utrecht.
- De Smet H. (2005). "A corpus of Late Modern English texts", *ICAME Journal* 29, 69–82.
- Diller H.-J., De Smet H. & Tyrkkö J. (2010). "A European Database of Descriptors of Electronic Texts", *The European English Messenger* 19(2), 29–35.
- Disney S. (2012). "The effect of error in the OED on grammaticalisation research", Presented at *New Reflections on Grammaticalisation V*, Edinburgh.
- Douglas F. M. (2003). "The Scottish Corpus of Texts and Speech: Problems of Corpus Design", *Literary and Linguistic Computing* 18(1), 23–37.
- Ekberg L. (1998). "Regeltillämpning Kontra Lexikonkunskap i Svenskan Hos Invandrarbarn i Malmö", in J. Møller, P. Quist, A. Holmen & J. N. Jørgensen (eds.) *Nordiske sprog som andetsprog*. København: Danmarks Lærerhøjskole, 247–263.
- Faarlund J. T. & Trudgill P. (1999). "Pseudo-coordination in English: the "try and" problem", *Zeitschrift Für Anglistik Und Amerikanistik* 47(3), 210–213.
- Flach S. (2014). "Solving rare phenomena problems with web-data? Quasi-serial verb constructions in English", Presented at *Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft (DGfS) 36* (Theme Section "Web Data as a Challenge for Theoretical Linguistics and Corpus Design"), Universität Marburg.
- Flach S. (2015). "Let's go look at usage: A constructional approach to formal constraints on go-VERB", *Yearbook of the German Cognitive Linguistics Association* 3, 231–251.
- Flach S. (2017a). "Idiomatic singleton or prototype? A productivity analysis of be-ADJ-and-V", *Yearbook of the German Cognitive Linguistics Association*, 5(1).
- Flach S. (2017b). *Serial verb constructions in English: A usage-based approach* (Ph.D. dissertation). Freie Universität Berlin.
- Fletcher P. & Garman M. (1988). "Normal language development and language impairment: Syntax and beyond", *Clinical Linguistics & Phonetics* 2(2), 97–113.
- Gachelin J.-M. (1991). "Transitivity and intransitivity in dialects of South-west England", in P. Trudgill & J. K. Chambers (eds.) *Dialects of English: Studies in Grammatical Variation*. New York: Longman, 218–228.
- Gillam R. B. & Pearson N. A. (2004). *Test of Narrative Language*. Austin, TX: Pro-Ed.
- Gries S. T. & Stefanowitsch A. (2004). "Extending collostructional analysis: A corpus-based perspective on 'alternations.'", *International Journal of Corpus Linguistics* 9(1), 97–129.

- Hansen K. D. (2003). *Brugen af maskulinum i færøsk*. Göteborg: Institutionen för svenska språket, Göteborgs universitet.
- Hansen K. D. (2005). "FTS – Føroyskt TekstaSavn / færøsk tekstkorpus", In H. Holmboe (ed.) *Nordisk Sprogteknologi [Nordic Language Technology] 2004: Aarvog for Nordisk Sprogteknologisk Forskningsprogram 2000-2004*. København: Museum Tusculanums Forlag, 47–50.
- Heycock C. & Petersen H. P. (2012). "Pseudo-coordination in Faroese", In K. Braunmueller & C. Gabriel (eds.) *Multilingual Individuals and Multilingual Societies*. Hamburg: John Benjamins, 259–280.
- Hommerberg C. (2003). *Pseudo-Coordination or to-construction: A corpus study of the "try and-problem."* Unpublished term paper, University of Växjö, English Department.
- Hommerberg C. & Tottie G. (2007). "Try to or try and? Verb complementation in British and American English", *ICAME Journal* 31, 45–64.
- Jaeggli O. A. & Hyams N. M. (1993). "On the independence and interdependence of syntactic and morphological properties: English aspectual *come* and *go*", *Natural Language & Linguistic Theory* 11(2), 313–346.
- Johannessen J. B., Priestley J., Hagen K., Åfarli T. A. & Vangsnes Ø. A. (2009). The Nordic Dialect Corpus — an advanced research tool. *Proceedings of the 17th Nordic Conference of Computational Linguistics (NODALIDA) 2009*, 73–80.
- Johnson M. G. (1986). *A computer-based approach to the analysis of child language data* (Ph.D dissertation). University of Reading, Reading, UK.
- Klemola J. (2000). "The Origins of the Northern Subject Rule: A Case of Early Contact?", In H. L. C. Tristram (ed.) *The Celtic Englishes II*. Heidelberg: Universitätsverlag C. Winter, 329–346.
- Kume Y. (2009). "On double verb constructions in English: with special reference to grammaticalization", *English Linguistics* 26(1), 132–149.
- Lieven E., Salomo D. & Tomasello M. (2009). "Two-year-old children's production of multiword utterances: A usage-based analysis", *Cognitive Linguistics* 20(3), 481–507.
- Lind Å. (1983). "The variant forms *try and/try to*", *English Studies* 5, 550–563.
- MacWhinney B. (2000). *The CHILDES Project: Tools for analyzing talk* (Third edition). Mahwah, NJ: Lawrence Erlbaum Associates. <http://childes.psy.cmu.edu/>
- Maia J. de C. (2012). Complementation patterns of the verb *try*. *Revista Virtual Dos Estudantes de Letras (ReVeLe)*, 4.
- Meurman-Solin A. (1995). "A New Tool: The Helsinki Corpus of Older Scots (1450-1700)", *ICAME Journal* 19, 49–62.
- Michel J.-B., Shen Y. K., Aiden A. P., Veres A., Gray M. K., Pickett J. P., Aiden E. L. (2011). "Quantitative Analysis of Culture Using Millions of Digitized Books", *Science* 331(6014), 176–182.
- Newman J. & Rice S. (2008). "Asymmetry in English multi-verb sequences: A corpus-based approach", in B. Lewandowska-Tomaszczyk (ed.) *Asymmetric Events*. Amsterdam; Philadelphia: John Benjamins Publishing Company, 3–23.
- Nicoll S. (2009). "Go-and-V, come-and-V, go-V and come-V: A corpus-based account of deictic movement verb constructions", *English Text Construction* 2(2), 185–208.

- Nordquist D. (1998). "English *try and*: A discourse analysis", in C. Berkenfield, D. Nordquist & A. Grieve-Smith (eds.) *Proceedings of the First Annual High Desert Linguistics Society Conference*. Albuquerque, NM: High Desert Linguistics Society, 119–128.
- Pollock J.-Y. (1994). "Checking Theory and Bare Verbs", In G. Cinque & R. S. Kayne (eds.) *Paths towards universal grammar: studies in honor of Richard S. Kayne*. Washington, D.C.: Georgetown University Press, 293–310.
- Pullum G. K. (1990). "Constraints on intransitive quasi-serial verb constructions in modern colloquial English", *Ohio State University Working Papers in Linguistics* 39, 218–239.
- Purves D. (2002). *A Scots grammar: Scots grammar and usage* (Revised and extended edition). Edinburgh: Saltire Society.
- Rasmark T. (2005). "Ett system som heter DUGA eller Føroysk TekstaSavn", In H. Holmboe (ed.) *Nordisk Sprogteknologi [Nordic Language Technology] 2004: Aarbog for Nordisk Sprogteknologisk Forskningsprogram 2000-2004*. København: Museum Tusculanums Forlag, 51–55.
- Rohdenburg G. (2003). "Cognitive complexity and horror aequi as factors determining the use of interrogative clause linkers in English", in G. Rohdenburg & B. Mondorf (eds.) *Determinants of Grammatical Variation in English*. Berlin: Walter de Gruyter, 205–249.
- Rosen A. (2014). *Grammatical variation and change in Jersey English*. Amsterdam: John Benjamins.
- Ross D. (2013). "Dialectal variation and diachronic development of *try*-complementation", *Studies in the Linguistic Sciences: Illinois Working Papers* 38, 108–147.
- Ross D. (2014). "The importance of exhaustive description in measuring linguistic complexity: The case of English *try and* pseudocoordination", in F. J. Newmeyer & L. B. Preston (eds.) *Measuring Grammatical Complexity*. Oxford: Oxford University Press, 202–216.
- Ross D. (2015). "What can Faroese pseudocoordination tell us about English inflection?", *LSO Working Papers in Linguistics (University of Wisconsin-Madison)* 10, 74–91.
- Ross D. (Forthcoming). *Pseudocoordination, serial verb constructions and multi-verb predicates: The relationship between form and structure* (Ph.D. dissertation). University of Illinois at Urbana-Champaign, Urbana, IL.
- Ruuska S. (2013). *Englishes Online: A comparison of the varieties of English used in blogs* (BA Thesis). Linnaeus University, Växjö, Sweden.
- <http://urn.kb.se/resolve?urn=urn:nbn:se:lnu:diva-27491>
- Santin-Guettier A.-M. (2005). "Des constructions *try to P* et *try and P* en anglais contemporain", in G. Girard-Gillet & L. Roux (eds.) *Parcours linguistiques: domaine anglais*. Saint-Étienne: Publications de l'Université de Saint-Étienne, 167–179.
- Snyder W. (2007). *Child language: the parametric approach*. Oxford: Oxford University Press.
- Snyder W. (2008). "Children's grammatical conservatism: Implications for syntactic theory", In T. Sano, M. Endo, M. Isobe, K. Otaki, K. Sugisaki & T. Suzuki (eds.) *An enterprise in the cognitive science of language: a festschrift for Yukio Otsu*. Tokyo: Hituzi Syobo, 41–51.
- Snyder W. (2011). "Children's grammatical conservatism: Implications for syntactic theory", in N. Danis, K. Mesh & H. Sung (eds.) *Proceedings of the 35th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, 1–20.
- Southwood F. (2007). *Specific language impairment in Afrikaans: providing a minimalist account for problems with grammatical features and word order*. LOT, Utrecht.

Stefanowitsch A. & Gries S. T. (2003). "Collostructions: Investigating the interaction of words and constructions", *International Journal of Corpus Linguistics* 8(2), 209–243.

Sugisaki K. & Snyder W. (2013). "Children's Grammatical Conservatism: New evidence", in M. Becker, J. Grinstead & J. Rothman (eds.) *Language Acquisition and Language Disorders*. Amsterdam: John Benjamins, 291–308.

Tottie G. (2009). "How different are American and British English grammar? And how are they different?", in G. Rohdenburg & J. Schlüter (eds.) *One language, two grammars?: differences between British and American English*. Cambridge: Cambridge University Press, 341–363.

Tottie G. (2012). "On the History of try with Verbal Complements", in S. Chevalier & T. Honegger (eds.) *Words, Words, Words: Philology and Beyond: Festschrift for Andreas Fischer on the Occasion of his 65th Birthday*. Tübingen: Narr Francke Attempto, 199–214.

Trudgill P. (2011). *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*. Oxford: Oxford University Press.

Uusi-Mäkelä T. (2013). *A Diachronic Study on the Complementation of the Verb Try* (MA Thesis). University of Tampere, Tampere, Finland.

<http://urn.fi/urn:nbn:fi:uta-1-23763>

Vosberg U. (2006). *Die Grosse Komplementverschiebung: aussersemantische Einflüsse auf die Entwicklung satzwertiger Ergänzungen im Neuenglischen*. Tübingen: Gunter Narr Verlag.

Wulff S. (2006). "Go-V and go-and-V in English: A case of constructional synonymy?", in S. T. Gries & A. Stefanowitsch (eds.) *Corpora in cognitive linguistics: corpus-based approaches to syntax and lexis*. Berlin: de Gruyter, 101–125.

Zwicky A. M. (1969). "Phonological constraints in syntactic descriptions", *CUHK Papers in Linguistics* 1(3), 411–463.

Zwicky A. M. (2003). *Go look at the modern language to test hypotheses about the past*. Abstract, Stanford University.

<http://web.stanford.edu/~zwicky/lisaabst.qsv.pdf>

NOTES

1. It must be noted that the BFC, and corresponding grammaticality, applies only to the pseudocoordination control-reading of *try and* (effectively synonymous with *try to*), in contrast to literal coordination as in 'They both tried (their best) and then succeeded.' Corresponding grammatical sentences with the normal coordination reading are set aside in the remainder of this article. See Ross (2013, 2014, forthcoming) and references therein for background on *try-and-V*.

2. Compare *remember* as the 380th most frequent word in COCA, but with under .5 instances of *remember and V* per million words, most of which are not pseudocoordination (e.g., literal coordination such as *remember and forget* or *remember and be remembered*); or *be sure* (collocation frequency: 28 per million words) with about 0.2 instances of *be sure and V* per million words. Therefore *try* is 20–50 times more likely to provide sufficient data to answer the research questions and also the most prototypical instantiation of the construction, thus likely to lead any innovation. On *be sure*, see also Flach (2017a).

3. See Disney (2012) on the problems of using the OED as a corpus.

4. Tottie (2012) included all of the available books at the time of writing (approximately 25,000, from 1475-1700; cf. Tottie 2012: 209).
5. No earlier verified instances of either construction have been found, whether in EEBO or any other source, by Tottie, Ross or any other author as far as I am aware. See those papers for discussion.
6. Smoothing is set to 50 years; data from Google Ngrams is presented as relative frequencies of N-grams, in this case tri-grams, and as such can only be interpreted relatively to other N-grams in the corpus.
7. For example, it could be averaged out over the period from 1810-1829, for which the frequency would be measured as .12 as opposed to .85 for the 1810s, representing less than one-sixth the amount of corpus text for the 1820s for which zero tokens were found.
8. Uusi-Mäkelä considered only two sections of the smaller original corpus and a subset of tokens from the latter section, but for all uses of the verb *try*, not just those with verbal complements.
9. As in the Google Ngrams data and elsewhere in the current study unless otherwise mentioned, only uninflected instances are considered when making frequency comparisons between the two constructions due to the BFC restricting *try-and-V* to uninflected verbs. Exceptions to the BFC are mentioned whenever they are identified and absent elsewhere. Inflected contemporaneous usage of *try-to-V* is also generally found but only discussed when relevant.
10. See also Brook & Tagliamonte (2016: 310–311, 315–316) for discussion of shifting usage frequencies today relative to speaker age in Canadian and British sociolinguistic interview corpora.
11. Now we can return to the analysis by Rohdenburg (2003) and Vosberg (2006) —mentioned, and rejected, above— because it focused on this period of development for *try-and-V*. Indeed, the frequency of usage increased rapidly, which may explain why the construction appeared to emerge then. Additionally, it was at this point that the construction took on distinctive grammatical properties (i.e. the BFC) of its own, correspondingly interesting also to the apparent development of *horror aequi*, or the avoidance of the equal, as described by Rohdenburg and Vosberg. That is, *try-and-V* is more frequent following *to* in infinitives, while *try-to-V* is more frequent following *and*. Although not the origin of the construction, it does correspond to the non-finite origin identified above, and also relates to the construction settling into the English grammatical and usage system and increasing in frequency in the 1800s.
12. Contra the speculation in Brook & Tagliamonte (2016: 317), the imperative appears to have been an early development in the construction. Beyond the one imperatival example from EEBO mentioned above, early examples are hard to find probably because imperatives are generally infrequent in early written texts and because the construction was low-frequency (no examples have been identified in Shakespeare's works for example), but after dense datasets are available in corpora, imperatival examples can be found, though they are still rare in writing:
 - (i) ...my child is gone — [t]hat loss is irretrievable: but, in the name of God, try and avert the ravages of a cruel disease from the rest of my family. (1803, Hansard corpus)
 - (ii) Try and see if you can avoid contradictions.
(Wesleyan-Methodist Magazine 1813, Google Ngrams)
13. Having consulted a wider range of corpora, these examples (17-18) are earlier than the corresponding earliest examples in Ross (2013: 120–121), although the general timeline is still accurate. Present-tense licensing *be* was initially very rare, with relatively consistent usage not found until the late 1900s (Ross 2013: 121). To fully understand the timeline of development for this aspect of the construction and BFC, substantially more data would be required.
14. Residual effects of the construction's origin can still be observed, however. Lind (1983), Hommerberg & Tottie (2007: 50–51) and others find varying frequency distributions based on syntactic context, even within the forms not restricted by the BFC. In modern usage, the most frequent contexts for *try-and-V* are still infinitives and imperatives, reflecting its earliest usage.

15. On grammatical differences in general, see especially Lind (1983). A particularly early source is Bryant (1962: 220-221), discussing frequency of *try and* in spoken vs. written English, summarized from informal (and unpublished) corpus 'studies' submitted by contributors to the volume such as by counting instances in newspapers; the BFC is not mentioned. Several other corpus studies such as Nordquist (1998) and Santin-Guettier (2005) investigate potential semantic differences between *try-and-V* and *try-to-V*, but as concluded by Hommerberg & Tottie (2007: 59) and Ross (2013: 110) and others, the difference between the two constructions is not semantic but in usage: *try-and-V* is colloquial and emphatic, potentially giving rise to conversational implicatures (even contradictory attitudes such as encouraging the attempt or disbelief). And this is also supported by the fact that despite the very different frequencies of *try-and-V* and *try-to-V* in British and American English, the sum of tokens per million words is similar in both dialects: see for example, the data in Hommerberg & Tottie (2007: 48) or data from COCA and BNC used as endpoints for Figures 3-4 in Section 2.2, although see Tottie (2009: 346-347,362) who observes a statistically significant difference of about 20 pmw (or 10%) more for British versus American English for the total frequency of both *try+verb* constructions and wonders what different discourse or lexical effects could account for that.

16. Based on unpublished work by Hommerberg (2003).

17. Davies (2014: 32-33) and Davies & Fuchs (2015: 14) also mention the frequencies of *try-and-V* in GloWbE but as a question of prescriptivism, disconnected from other research on the construction, and without looking at the BFC.

18. The dialects are identified in the GloWbE corpus as follows: AU = Australia; BD = Bangladesh; CA = Canada; GB = Great Britain; GH = Ghana; HK = Hong Kong; IE = Ireland; IN = India; JM = Jamaica; KE = Kenya; LK = Sri Lanka; MY = Malaysia; NG = Nigeria; NZ = New Zealand; PH = Phillipines; PK = Pakistan; SG = Singapore; TZ = Tanzania; US = United States; ZA = South Africa. Within each dialect, the percentage values are based on 5,000-76,000 tokens of *try*.

19. No examples were found for another possible type with bare *try* but an inflected second verb (e.g., **He try and goes*), aside from typos unmotivated by syntactic context (e.g., *So don't try and stops me!*).

20. Extremely rare exceptions to the BFC are also found in other corpora. For example, among over 12,000 examples of *try-and-V* in the Hansard corpus (see Section 2.2), there are only two exceptions:

Some of the money... might as readily have been spent..., in trying and find new work for men who are out of work, to increase work and wages...

(Thomas Sexton 1938, Hansard corpus)

... when one comes to consider this matter in the light of the considerations which I have put forward, and tries and thinks of some really effective practical proposal...

(Sir Kingsley Wood 1941, Hansard corpus).

21. Similarly, some speakers of African American English may accept present-tense habitual usage of *be sure and* (Ross 2013:124), thus showing the same BFC is operative for that verb despite being necessarily non-finite due to its paradigm in standard English.

22. But see Purves (2002: 65-66) for examples of imperativial pseudocoordination in Scots with several verbs including *be shuir* and *mynd*, though no examples specifically of *try* are given.

23. Because their data consisted of single-session sociolinguistic interviews, it is possible that speakers would vary more across multiple sessions or discourse contexts, which would be an important question to pursue in future research.

24. Or individual variation could be more incremental, along the lines discussed by Pullum (1990) for the *go-get* construction, such as whether present-tense *be* is accepted by different speakers. Additionally, if relaxed, would the BFC give way to two parallel inflection forms, or an inflected first verb and infinitive second verb?

25. Where the number of children and number of tokens varies (due to multiple tokens being produced by one or more children), the more conservative figure (number of children) is used for statistical tests.
26. These normally developing children were the control group in a study that also included 250 disordered children, whose speech is not analyzed here. However, only one instance of *try-and-V* is found in this corpus, perhaps suggesting that *try-and-V* may be difficult for children with language disorders to acquire, but compared to only 7 tokens produced for *try-to-V* this evidence is weak. The only relevant study along these lines is for a different type of pseudocoordination construction in Afrikaans (with posture verbs, often corresponding to progressive aspect) by Southwood (2007), who found that children with Specific Language Impairment may have some difficulty on grammaticality judgment tasks compared to normally developing children, but the results for the pseudocoordination category in particular were not statistically significant. This line of inquiry for English *try-and-V* must be set aside for future research.
27. The corpus is not balanced by number of children per age nor length of the narrative told by each child, so this table does not display a representative distribution.
28. Additionally, one agreement error was found: *He try to get*.
29. The high frequency of this form (as well as *trying to*) is due to the narrative format of the corpus, with primarily past-tense stories.
30. However, even if we adjust the data to a more reasonable tense distribution, the result may still be significant. For example, based on the spoken section of COCA which has about 2.4 inflected tokens of *try-to-V* for every bare token, there would be corresponding the 14 bare tokens approximately 34 inflected tokens in the more normal discourse situations represented by COCA. If we substituted 34 for 270 in the table above, the result would still be statistically significant ($p < .05$).
31. The literal coordination reading can be paraphrased as 'The girl gives it a try (taking him over there), and tugs on him'. The *try-and-V* reading is 'The girl tries to tug him over there.' It is not entirely clear from context whether she has yet succeeded in tugging him or not.
32. This construction has been analyzed by a number of linguists (Zwicky 1969, Carden & Pesetsky 1977, Pullum 1990; Jaeggli & Hyams 1993, Pollock 1994, Bjorkman 2016, and Ross forthcoming, among others). See especially Flach (2014, 2015, 2017b), whose work discusses this construction as a low-frequency phenomenon in corpus research. The diachrony of the construction is also similar in that there is an alternation between *go-V* and *go-and-V*, with *go-V* restricted to uninflected usage, while *go-and-V* can be fully inflected (in contrast to *try-and-V*). The origin of *go-V* and *go-and-V* and whether *go-V* is derived from *go-and-V* has been the subject of debate (Zwicky 2003; Nicolle 2009; Kume 2009).
33. Only the first two collexemes, marked in bold, were found to be statistically significant. One-token results are not included in the list.
34. The most frequent collocations of *try-and-V* identified by Brook & Tagliamonte (2016: 312) include *get*, *find*, *make*, *keep*, *go*, *take*, *put*, *sell*, *stay*, *teach*, *help*, *stop*, and others either less frequent or substantially more attracted to *try-to-V*. A similar list of most frequent verbs used with *try-and-V*, but not in contrast to usage with *try-to-V*, is given by Newman & Rice (2008: 20): *get*, *do*, *find*, *make*, *go*, *put*, *see*, *learn*, *play*, *say*, *take*. Maia (2012: 17) also provides a similar list for *try-and-V* including especially *get* but also *do*, *keep*, and *make*, but these are also found to be among the most frequent with *try-to-V*, probably because of a problematic methodology: Maia included inflected instances of *try*, which are of course heavily biased toward the *try-to-V* construction, which is used even with frequent collocates for *try-and-V* when that construction would be ungrammatical. Thus the list from Gries & Stefanowitsch (2004) is the most representative of differences between these constructions, even with limited statistical significance.
35. Thomas was recorded one hour per day, five days per week at age 2, for the whole year, so using these tokens as a proxy for his first usage of the construction is not unreasonable.

36. Abbreviations: CIT: citation form or unclear grammatical context/fragment (e.g. *I can do this: try and put it there.*); COORD: following *and* (*and try and*); IMP: imperative; INF: bare infinitive complement (e.g. *will try and*); PRES: bare present tense (e.g. *I try and*); TO: to-infinitive complement (*to try and*).

37. Note also the data from GloWbE in Section 3, where the the lowest usage of *try-and-V* is found in many of the dialects that include high numbers of non-native speakers (Bangladesh, Philippines, Pakistan, etc.). Similarly, Ekberg (1998) found that in Swedish non-native speakers used aspectual pseudocoordination constructions (with motion or posture verbs) significantly less often than native speakers.

38. Furthermore, the few corpora representing dialectal variation in Faroese have very small sample sizes for each dialect, such as the Nordic Dialect Corpus (Johannessen et al. 2009) with only six tokens for *royna* 'try' and none for *royna og* 'try and', leaving only general corpora for Faroese as feasible candidates at this time for research on low-frequency syntactic phenomena.

ABSTRACTS

Small corpora remain an important resource for addressing questions requiring specialized data not found in large general purpose corpora. But do these small corpora have sufficient data to analyze low-frequency phenomena? This article presents a case study of the *try-and-V* pseudocoordination construction in English, which has an unusual morphosyntactic restriction: neither verb can be inflected. Diachronic corpora elucidate its origin as reanalysis from ambiguous usage of non-finite verbs in the late 1500s. Dialectal corpora reveal variation in frequency but not grammatical properties. And child language corpora reveal productive usage even by young children. These results are considered successful and the outlook is optimistic for addressing low-frequency phenomena beyond contemporary, standard English, especially as more, and more diverse, corpora become available.

Les petits corpus demeurent une ressource importante pour répondre aux questions nécessitant des données spécialisées qui ne se retrouvent pas dans les grands corpus à usage général. Mais ces petits corpus ont-ils des données suffisantes pour analyser les phénomènes de basse fréquence ? Cet article présente une étude de cas de la construction de pseudocoordination *try-and-V* (litt. « essayer et Verbe ») en anglais, qui présente une restriction morphosyntaxique rare : aucun verbe ne peut être conjugué. Les corpus diachroniques éclairent son origine par une réanalyse de l'usage ambigu de verbes non finis à la fin des années 1500. Les corpus dialectaux révèlent une variation de fréquence mais pas des propriétés grammaticales. Les corpus de la langue de l'enfant révèlent un usage productif même par les jeunes enfants. Ces résultats sont considérés comme un succès et la perspective d'un traitement des phénomènes de basse fréquence au-delà de l'anglais standard et contemporain est optimiste, surtout puisque davantage de corpus plus diversifiés deviennent disponibles.

INDEX

Mots-clés: petits corpus, pseudocoordination, phénomènes de basse fréquence, développement diachronique, variation dialectale, acquisition

Keywords: Small corpora, pseudocoordination, low-frequency phenomena, diachronic development, dialectal variation, acquisition

AUTHOR

DANIEL ROSS

University of Illinois, Urbana-Champaign