To Split or to Not Split: The Split Infinitive Past and Present

Moisés D. Perales-Escudero[1]

Abstract

This article reviews the history of the proscription of the English split infinitive and presents a corpus-based investigation of its present usage and distribution across registers in American English. Using archival research, the article traces the history of the proscription and offers historical examples of its use. Contrary to popular belief, the historical review of prescriptive sources reveals not a Latin origin for the proscription but a German one. Thus, an ideology of Teutonic kinship seems to have at least partially driven the proscription of the split infinitive. Latin-based proscriptions seemed not to have existed in written form, or if they did, they did not survive. The Corpus of Contemporary American English (COCA) was used to investigate the presence of multiword lexical items including the split infinitive. Combinations of "*to* + adverb" were identified in the different COCA registers. Distinct bigrams and trigrams emerge in the different registers of the COCA, but it does not seem to be the case that oral registers, where the force of the proscription would be less strong, influence the emergence of split infinitive patterns in written registers. The results are discussed within the framework of prescriptivist ideology, grammaticalization, and idiomaticity.

Keywords

prescriptivism, grammar, corpus linguistics, idiomaticity

To boldly split one's infinitives may seem, to some, a rather innocent and trivial act rather than an instance of grammatical audacity or mischief. Yet there are others who, like U.S. Supreme Court Chief Justice John Roberts, still seem to cringe at the thought

[1]University of Michigan, Ann Arbor, MI, USA

Corresponding Author:
Moisés D. Perales-Escudero, University of Michigan, 435 S. State St., 3187 Angel Hall,
Ann Arbor, MI 48109-1003
Email: moisesd@umich.edu

of placing an adverb between the main verb and other components of a verb phrase. Thus, instead of having Barack Obama "solemnly swear that I will faithfully execute the office of President of the United States," Mr. Roberts had him "solemnly swear that I will execute the office of the United States faithfully." As Steven Pinker notes in "Oaf of Office" (Pinker 2009), Mr. Roberts's mangling of one of the best-known lines in the U.S. Constitution may be the result of the Chief Justice's adherence to "the prohibition against 'split verbs'" of which the split infinitive is an instance. Pinker (2009) identifies this prohibition as one of the main fetishes created and promoted by prescriptive grammarians of the English language. Indeed, there is evidence that few other prescriptive rules have occupied the attention of prescriptive grammarians and speakers of English to the extent that the split infinitive has. For example, Bailey (2006) reports finding over 87,000 Web sites where the topic is discussed. At the time of writing this article, a search for *split infinitive* using the popular search engine Google produces approximately 126,000 hits. By contrast, a search for the phrase *the dangling modifier* produces only approximately 76,400 hits. The plethora of sites devoted to the topic suggests that the split infinitive might be an object of linguistic insecurity and anxiety as well as interest, more so than other grammatical structures deemed to be "incorrect."

In stark contrast to the strong popular and prescriptivist interest in the subject, there is a paucity of published empirical research examining the split infinitive from a descriptive viewpoint. This shortage of research may be from a growing scholarly consensus that the issue regarding the correctness of the split infinitive has been settled. Oxford, after all, lifted the ban on the split infinitive in 1998 (s.v. "infinitive"). For example, the *New York Public Library Writers' Guide to Style and Usage* (1994:583) states, "Today, only the most rigid grammarians do not allow split infinitives at least once in a while." Many other sources agree that the split infinitive is not the capital offense that many English teachers and prescriptive grammarians think it is (Howard 1997; Partridge 1995; Peters 2004).

Nevertheless, there is evidence that the influence of the prescriptivist ideology surrounding the split infinitive, if weakened, is still strong. For example, the influence of the proscription is such that it manages to, not without irony, split the American Heritage Usage Panel in half when deciding on the acceptability of the split infinitive in the sentence, "The move allowed the company to legally pay the employees severance payments that in some cases exceeded $30,000" (s.v. "split infinitive," def. 59). Echoing the sentiments of those members of the American Heritage Usage Panel who condemn the structure, the *Webster's Dictionary of English Usage* (1989:867) cautions that "all the evidence points towards the reality of the feeling that it is 'wrong' to split infinitives."

It is then relevant to wonder how this feeling of wrongness exactly came into being, particularly in light of the fact that the structure has long been found in English prose (Curme 1931; Visser 1984). And given the prevalence of the structure as well as the ambivalence about it, could it be that specific collocational patterns of usage have emerged for certain "*to* + adverb + verb" combinations that make such patterns more

acceptable—more standard?—than the nonsplit alternative? Part of the answer to the first question is that the nonsplit infinitive has coexisted with the split infinitive for centuries. As Milroy and Milroy (1991:22) suggest, the coexistence of two forms for the same grammatical meaning of function is precisely what the process of standardization stands against: "standardisation does not tolerate variability." Crucially, standardization requires agents, language guardians, self-appointed and otherwise, who take it on themselves to ensure that one or another form should fall out of usage by casting it as substandard or deviant. While prescriptivist activity is usually dismissed by linguists, the Milroys' assertion that "a major task of sociolinguists is to explain why linguistic differences that are essentially arbitrary are assigned social values" (Milroy & Milroy 1991:19) suggests that it is important for linguists to analyze the ideologies driving particular prescriptions of specific forms.

For the split infinitive, both the documents that proscribe it and the ideologies underlying those proscriptions, to use Hall's (1882) fitting term, remain somewhat unclear. For example, Myers (2002) attributes the proscription to Bishop Lowth and his insistence on casting English grammar into the mold of Latin. More generally, Williams (2007) calls the proscription an invention of eighteenth-century grammarians with similar Latinate inclinations. This sentiment is echoed by the *American Heritage Dictionary of the English Language* when attempting to explain the rationale for banning the split infinitive in the following terms: "The only rationale for condemning the construction is based on a false analogy with Latin" (59). Yet Bailey (2006) considers what I call "the Latin origin hypothesis" to be a myth, part of the folklore of linguistics and, contra Myers and Williams, asserts that the proscription has its origin in the mid-nineteenth century; that is, not in Lowth, Murray, or any of the other eighteenth-century grammars. These discrepancies suggest that a thorough analysis of proscription's documentary sources and the ideologies behind them could usefully be added to the literature as an initial articulation of an answer to the first question above regarding the origins of this proscription. Furthermore, an analysis of the contemporary variability of the split infinitive across registers could offer some answers to the second question regarding the potential emergence of collocational patterns involving the split infinitive.

Scholarly discussions of the split infinitive have not, to the best of my knowledge, considered the possibility that the acceptability of some of the patterns involving the split infinitive that are presented as satisfactory or nonoffensive (e.g., "to really" in Strunk & White 1979) might be the result of their emergence as preconstructed phrases, "chunks," or lexical bundles. This article offers a corpus-based exploration of the question of whether any idiomatic patterns involving the split infinitive exist in contemporary spoken and written registers of American English. The results suggest that some such patterns, particularly "to better understand," seem to be emerging as acceptable patterns with strong register associations. As a framework for corpus-based findings, I also locate and analyze the proscriptive sources identified in the literature to better characterize (1) the origin of the proscription and (2) the ideologies of the language guardians who offered them. A historical analysis reveals that a preference

for the Teutonic nature of English, to the detriment of its Latin origins, is one such ideology. The findings from the analysis of the historical proscriptive sources are then contrasted with the current ideologies permeating the advice offered by contemporary usage guides. Such advice overwhelmingly relies on decontextualized notions of clarity and elegance and tends to be ambivalent about the target structure, despite the sources' quasi-uniform acknowledgment of the historical presence of the split infinitive in the English language. Within both the historical and the corpus-based investigation, I incorporate an analysis of *not* as the splitting element. Curme (1931) notes that the force of the proscription is stronger when *not* is the splitting element, but no empirical data exist to support his assertion.

A Brief History of the Split Infinitive and Its Proscription

The split infinitive has occurred in the English language for a long time, both with an adverb and with the negative particle *not* as the splitting elements. Curme (1931) presents the following examples from the fourteenth century. Notice that *not* is the splitting element in the second example:

(1)   He louied Þe lasse auÞer to lenge lye or to longe sitte

"He did not like to either lie or sit long" (Sir Gawayne and the Greene Knight, ll. 87-88, quoted in Curme 1931:460)

(2)   It is good to not ete flesich and to not drinke wyn

"It is good to not eat flesh and to not drink wine" (Wycliff, Romans, XIV, 21, Purvey's ed., A.D. 1388, quoted in Curme 1931:460)

In the fifteenth century, Richard Pecock, the English theologian and one of the first authors to have used English for theological writing, used *to not* freely:

(3)   Y schall swere to not discouere them

"I shall pledge myself to not inform on them" (id., The Folewer to the Donet, E.E.T.S., No. 164, p. 97 c. 1454, quoted in Curme 1931:460)

According to Fredericus Visser, the structure fell out of use in the written language from the sixteenth century until the last decades of the eighteenth century. In a similar vein, Fowler and Burchfield find evidence of the avoidance of the split infinitive in three different biographies of Sir Thomas More published in the sixteenth century. Here is an example:

(4) I am ready obediently to conforme my self to his graces' commandments.
(Roper, c. 1557, quoted in Fowler & Burchfield 1998:738)

However, not all scholars agree. Curme argues that the split infinitive can be found in the writing of distinguished sixteenth- and seventeenth-century writers such as Thomas Cromwell, William Tyndale, Samuel Pepys, Sir Philip Sydney, John Donne, Daniel Defoe, and Dr. Johnson. Curme, Fowler, and Burchfield agree that the construction became more common in the nineteenth and twentieth centuries. Writing in 1931, Curme found that over sixty renowned authors of literature, science, and political discourse publishing in those centuries used various forms of the split infinitive. The list includes the linguist Edward Sapir, the novelist Francis Scott Fitzgerald, the poet Lord Byron, and President Herbert Hoover. Curme provides an extensive number of sample sentences from those authors' works containing the split infinitive.

I now turn my attention to the origins of the proscription of the structure. As pointed out above, eighteenth-century grammarians (particularly Bishop Lowth) have widely been credited with inventing the proscription of the split infinitive. For example, Myers (2002:59) attributes the rule against splitting infinitives to Bishop Lowth: "It seems that Lowth reasoned that since it was grammatically impossible to split an infinitive in Latin, the language of learning and prestige in his day, then English shouldn't allow the split infinitive either."

As seen in Myers (2002), an associated claim is that such alleged proscription stemmed from an elitist ideology that sought to force English grammar into Latin forms for the sake of "improving" the former. Nevertheless, some scholars have located the sources of the proscription elsewhere. Bailey (2006), for one, rejects the Latin origin hypothesis in the following terms:

Many people who ought to know better—"authorities" on English—declare that the objection to separating *to* from the infinitive verb that follows is based on Latin (or some other language) where infinitives are single words. If some purist has made such a comparison, I can find no record of it. Henry Alford (The Queen's English [1866]) thinks the *to* and the verb are "inseparable" but he does not mention foreign languages (p. 227). This idea is part of the folklore of linguistics.

In support of Bailey, my own examination of the works of Lowth (1762/1967) and Murray (1792/1968), often cited as the two most influential eighteenth-century grammars of the English language, has rendered no evidence that they are the source of the proscription. Contra Myers (2002), I have found no condemnation of the split infinitive in Lowth; he does not discuss the structure at all. Murray (1792/1968:129), for his part, includes models of correctness where he seems to avoid the split infinitive, such as, "He must not expect to find study agreeable always." But nowhere does he explicitly condemn the use of the split infinitive.

There are, however, other potential sources for the proscription. According to Bailey, the first recorded prescription against the split infinitive appears in a 1834 anonymous letter to the editor of the *New England Magazine*, signed by "P." P explains that he or she does not know of any rules against the split infinitive but dislikes it because it is used only by "the uneducated"; and P volunteers to offer a rule, which he or she states as follows:

> The practice of separating the prefix of the infinitive mood from the verb, by the intervention of an adverb, is *not unfrequent among uneducated persons*; as "To fully understand it," instead of "to understand it fully," or "fully to understand it." This fault is not often found in print, except in *newspapers where the editors have not had the advantage of a good education*. I am not conscious that any rule has been heretofore in relation to this point: no treatise on grammar or rhetoric, within my knowledge, alludes to it. The practice, however, of not separating the particle from its verb, is so *general and uniform among good authors*, and the exceptions are so rare, that the rule which I am about to propose will, I believe prove to be as accurate as most rules, and may be found beneficial to inexperienced writers. It is this:—the particle TO, which comes before the verb in the infinitive mode, must not be separated from it by the intervention of an adverb *or any other word or phrase*; but the adverb should immediately precede the particle, or immediately follow the verb. (P, letter to the editor 1834)

Clearly, P's prescription operates within an ideology that proscribes the structure on the grounds of the social identity of its perceived users: the structure is wrong because it is the uneducated that use it. An appeal is also made to the prestigious usage of "good authors." This notion flies in the face of Curme's (1931) finding that several prestigious eighteenth- and nineteenth-century authors of English prose used the split infinitive. Notice also that P does not restrict the proscription to adverbs but includes any other word, which encompasses *not*.

On the other side of the Atlantic, and with a more authoritative editorial voice, the English editor Richard Taylor was perhaps the second person to proscribe the split infinitive in print. In the preface to John Horne Tooke's 1840 work *The Diversions of Purley*, Taylor discusses the changes in Old English and German leading to the present "*to* + verb" infinitive. As a corollary to his assertion that "in German, 'zu' is prefixt to a verb governed by another verb that precedes it" (Taylor 1840:xxx), he writes that "[s]ome writers of the present day have a disagreeable affectation of putting an adverb between 'to' and the infinitive" (Taylor 1840:xxx). It seems then that Taylor's proscription of the split infinitive is a case of trying to bring English closer to its Teutonic origins. Taylor's proscriptive rationale thus echoes calls to purify English via the removal of Greco-Latin words made in previous centuries by Thomas Wilson and John Cheke and in the nineteenth century by William Barnes.

As both Bailey and Partridge have noted, Henry Alford's (1863) *A Plea for the Queen's English* is another important source of the proscription of the split infinitive. In this influential work, Alford plainly states that the split infinitive is not part of the English language:

> Adverb between "to" and the infinitive.—A correspondent states as his own usage, and defends, the insertion of an adverb between the sign of the infinitive mood and the verb. He gives an instance, "to scientifically illustrate." But surely this is a practice entirely unknown to English speakers and writers. It seems to me that we regard the *to* of the infinitive as inseparable from its verb. And when we have already a choice between two forms of expression, "scientifically to illustrate" and "to illustrate scientifically," there seems to be no good reason for flying in the face of common usage. (Alford 1863:133)

Alford had obviously not paid much attention to the writing of the infinitive splitters cited by Curme (1931).

This short review has shown that there is no evidence in favor of the Latin origin hypothesis or the associated eighteenth-century hypothesis. If assertions proscribing the split infinitive on the grounds of enforcing Latinate standards were ever made in that or another century, they seem not to have been recorded or survived. Importantly, however, there is evidence that the proscription was justified through a comparison with another language: German. This is, to my knowledge, the first evidence tracing this proscription to an ideology of Teutonic kinship or purity.

The Split Infinitive in Contemporary Usage Guides

Howard (1997:341) offers the following piece of advice in the *Macmillan Good English Handbook*:

> If you don't want to upset anyone, you will avoid split infinitives. If you care more about writing good clear English, you will be prepared to fearlessly split an infinitive to allow words to fall in their natural place. But there's no need to split infinitives just for the sake of it: "to" and its following verb belong to each other, and should be separated only when good sense and the natural flow of words require it.

This excerpt provides a summative representation of the state of the argument in prescriptive resources regarding the "correctness" of the split infinitive. On one hand, some sources, such as the *Webster's Dictionary of English Usage* and the *Webster's New World Student Writing Handbook*, still advise against using the structure more or less explicitly. The author of the latter, for example, says, "Generally, experienced writers avoid split infinitives" (Sorenson 1997:579). On the other hand, there are

resources, such as the *New York Public Library Writers' Guide to Style and Usage*, that openly condone the use of the split infinitive. The prevalent position in most of the usage guides consulted in this research study, though, was one of ambivalence within the same entry, as seen in the following pieces of advice:

> Avoid splitting infinitives whenever possible, but do not suffer undue remorse if a split infinitive is unavoidable. (Fowler & Burchfield 1998:738)

> Don't split an infinitive if the result is an inelegant sentence. Do split infinitives to avoid awkward wording, to preserve a natural rhythm, and especially to achieve the intended emphasis and meaning. (Peters 2004:513)

Avoid the split infinitive wherever possible; but if it is the clearest and the most natural construction, use it boldly. The angels are on our side. (Partridge 1999:309)

> The construction should be avoided unless the writer wishes to place unusual stress on the adverb. (Strunk & White 1978:58)

> The split infinitive is another trick of rhetoric in which the ear must be quicker than the handbook. . . . A matter of ear. (Strunk & White 1979:78)

In all of the usage guides above, a recommendation against the structure is followed by a suggestion that it can be used under certain circumstances. In most of these entries, those circumstances are characterized in terms of elegance, natural rhythm, clarity, or "ear." These are all subjective categories that offer little by way of concrete evidence or accounts of usage patterns in which the structure is acceptable. Nevertheless, the use of specific adverbs in the examples of acceptable patterns, such as "really" in Strunk and White and in Peters below, suggests that these authors may have intuitive perceptions of emerging lexical patterns with the split infinitive:

> A single-word adverb runs in smoothly enough, especially an intensifier: He wanted to really talk to her. (Peters 2004:512)

David Crystal's (2006:126) is perhaps the only voice suggesting that there might be some kind of linguistic rationale for splitting infinitives that can be described objectively. He suggests that splitting an infinitive is part of native-like competence because doing so adjusts to the natural rhythm of the English language, which he characterizes as the "iambic pentameter, with strong (stressed) and weak (unstressed) syllables alternating." Thus, a sentence like "to boldly love" has a suprasegmental pattern that goes from weak to strong to weak to strong again. Crystal, however, does not offer any further examples to support his claim. Nor does he address the possibility that patterns involving the split infinitive may exist in the language that could further support, or undermine, his claim.

The Split Infinitive in Descriptive Grammars of English

Descriptive grammarians agree that there is no rational basis for the proscription of the split infinitive. For example, Huddleston and Pullum (2002:581) view the particle "to" in infinitive constructions as a subordinator, similar to "that" or "whether," and therefore conclude that

> in such a VP as "to love her" the immediate constituents are "to" and "love her," so that "to love" does not form a syntactic constituent, let alone a word. From a grammatical point of view, therefore, the adjunct in "to genuinely love her" does not split anything.

Quirk et al. (1985:496) note that the split infinitive is common in the speech of educated native speakers and add that "split infinitives are commonest with subjuncts of narrow orientation,"[1] particularly those with a focus orientation, as in "you ought to at least try" or "I'm going to really hurry."

Contemporary work in corpus linguistics suggests that strings of specific words, variously called lexical phrases, lexical bundles, or chunks, can occur together with higher than random frequency and are an important component of adult native language (Altenberg 1990). In particular, the existence of recurring three- and four-word combinations occurring more than ten or twenty times per million words, or lexical bundles, is well documented (Biber, Conrad, & Cortes 2004; Biber & Barbieri 2007).[2] Importantly, it has also been found that lexical bundles are unevenly distributed across registers in terms of frequency. For example, dependent-clause lexical bundles such as "if you look at" are more common in academic conversations, whereas NP/PP-based lexical bundles such as "shown in figure N" is more common in published articles (Biber, Conrad, & Cortes 2004:398). The possibility that lexical bundles involving the split infinitive may both exist in the English language and be distributed across registers has not been pursued by either prescriptive or descriptive sources. The following section reports a corpus-based study offering an initial characterization of some such patterns and their register variation.

The Corpus-Based Study
Method

The Corpus of Contemporary American English (COCA) was used for this study. The COCA is an online, 385-million-word corpus covering the following registers roughly equally: spoken (newscasts), fiction, popular magazines, newspapers, and academic texts. The texts included cover the years 1990 through 2008, with about 20 million words for each year. It was selected because of the facts that it is a free-access corpus, it is representative of a variety of registers, and it is large enough that results can be claimed to represent more or less general trends in American English. Word

searches were conducted using the online interface provided with the corpus by its compiler, Mark Davies.

This corpus-based study investigates whether any collocational patterns involving the split infinitive with adverbs and with "not" obtain in the COCA. With regard to adverbs, the COCA was searched for the pattern "*to* + adverb" or, in the corpus's conventions, "to [*r]." A list of the one hundred "*to* + adverb" combinations was obtained, including combinations such as "to just," "to really," "to further," "to now," "to ever," and so on. Each of the items in this list was then individually searched to determine its frequency of occurrence across registers with the purpose of selecting the most frequent items to conduct individual searches of their combinations with verbs. A cutoff point was arbitrarily set at five occurrences per every one million words in any given register. Items that did not meet this minimum frequency of occurrence were not included in the next stage of the study. The items included were "to just," "to really," "to actually," "to better," "to fully," and "to effectively." Examples of excluded items are "to only," "to nearly," "to always," "to completely," and "to significantly."

For each one of the remaining items, those occurring more than five times per one million words, the adverbs alone were searched with the goal of determining whether the register distributions of the adverbs, when matched with the register distribution of the split infinitive with those adverbs, revealed any significant patterns of register-motivated association. For example, the adverb *just* occurs 316.5 times per million words in spoken registers versus 60 times in academic registers. The adverb is then 5.27 times more common in spoken registers than in academic ones. It could then be expected that "to just" would also be roughly 5.27 times more common in spoken registers, but it actually is 22.9 times more common. If these two numbers are divided (22.9/5.27), a ratio of 4.34 is obtained. If the frequencies in both registers were equal, a ratio of 1 would obtain, which would mean that the frequency of the "*to* + adverb" patterns would be a function of the frequency of the adverb. So the higher the ratio, the less likely it is that the variation in frequency is due only to the distribution of adverbs; a higher ratio suggests a stronger, or higher than random, register-motivated association. This register association ratio was obtained for each of the six items included. The item with the highest register association ratio was "to better," and it was also a ratio that ran counter to expectations because "to better" is much more common in academic registers than it is in spoken ones despite the fact that the opposite is true for "better" alone. For this reason, the next stage of the study focused on "to better."

The next step was to search for verb collocates of "to better." The COCA allows for two kinds of collocational searches: MI-based collocations and frequency-based collocations. The MI, or mutual information score, expresses the extent to which frequency of co-occurrence is different from what could be expected based on the overall frequencies of the relevant words or sets of words in the corpus. Generally, MI scores higher than 3 indicate a strong semantic bond between words (Davies, *Corpus of Contemporary American English*). A problem with the MI is that it can be very high for low frequency words and/or low frequency collocates. For example, the verb

"stratify" occurs only once with "to better," which is to say that there is only one instance of "to better stratify" in the corpus. Yet, its MI is 10.94. Clearly, calling "to better stratify" a significant pattern based only on the MI would be misleading. Furthermore, MI-based tables in the COCA may leave out very frequent patterns, such as "to better serve," whose MI scores are not high enough to make the first one hundred, which is what the COCA displays. For those reasons, MI-focused collocational searches need to be complemented with frequency-focused collocational searches. And only items occurring more than ten times were selected for inclusion in MI-focused tables, which means that low-frequency items such as "stratify," "synchronize," and "visualize" were left out.

The COCA provides MI scores only when collocates are searched without specifying a grammatical category and when windows of more than two words are used on either side of the context search. This kind of search was conducted first. Then, a second, frequency-focused search was conducted using the label [*v] or "verb" immediately after "to better," without specifying a context window. This search produced collocates arranged by frequency. The first ten collocates were selected for inclusion in tables. The next step was to search for avoidance of the patterns found in both types of searches above. For example, it was necessary to determine whether the avoidance of "to better understand," as in "teachers need to understand the abilities of a DDT pupil better" (an actual clause from the COCA), was more frequent in terms of number of clauses showing such avoidance than the number of occurrences of "to better understand" as a string. To do this, the "*to* + verb" combinations found in the MI-oriented and frequency-oriented searches were searched as continuous strings with "better" specified as context within a one-word window to the left and a nine-word window to the right. The latter is the greatest number of words allowed for context windows by the COCA. Leaving a window of this length ensured that clauses such as the one about DDT pupils above would be captured. But it also meant that several clauses and sentences not showing avoidance of the split infinitive were also included in the counts, such as "they also seemed to understand better than their elders the ways in which Washington has changed" (also an actual sentence). For that reason, verb collocates occurring less than fifty times were selected to be analyzed by hand to get more precise figures for avoidance. The cutoff point of less than fifty times was set for convenience. An analysis of the alleged avoidance patterns for all verb collocates, particularly those occurring hundreds of times, would be beyond the scope of this study. The register distribution of strings found to have high frequencies as well as low avoidance, such as "to better understand" were then searched.

A search for the particle "not" as a splitting element involved a different method for two reasons. First, "to not" is a combination existing in opposition to "not to." This opposition suggests that search criteria must seek to account for the different frequencies of "to not" and "not to." Second, because it is not necessary to characterize the frequency of different splitting elements (there is only one), it becomes interesting to determine collocations on both the right and the left of "to not," such as "to not want"

(with "want" as a right-hand collocate) versus "want to not" (with "want" as a left-hand collocate). Once such collocations are found for "to not," the opposition with "not to" is functional again: the question is which is more common, either "to not want" or "not to want," or "a tendency to not" versus "a tendency not to" in the case of nonverbal collocates. These considerations provide the rationale for the steps below.

Both "to not" and "not to" were searched as word strings to find out which one was more common in which register. Then, "to not" was searched for potential collocates on both the left and the right. As indicated above, the potential existence of nonverbal collocates, as in "a tendency to not" leads to the specification of context as so that results would include words in all grammatical categories, not only verbs. A window of three words to the left was used to expand the possibilities of capturing collocates in phrases such as "a tendency in Americans to not" (an invented example). For similar reasons, a window of four words to the right was used to account for the possibility of there being two intervening adverbs between "to not" and potential verb collocates, as in "to not only and exclusively want" (an invented example). This was, of course, a MI-focused search. Also, the tables produced by the interface did not group lemmas together (i.e., there were separate entries for "try" and "tried"). Lemmas were grouped together and their frequencies were added to make both MI-based and frequency-based tables. Then, frequency-focused searches were conducted to complement the results of the MI-focused search. Having obtained patterns, instances of avoidance were sought. For example "tend to not" was found to be an interesting pattern, and then "tend not to" was searched to compare frequencies.

Results

Table 1 shows the "*to* + adverb" combinations that occur more than five times per million words in any given register, as well as the normalized counts. As noted above, these were selected out of the initial one hundred most frequent such combinations for the next stage of the analysis.

Some clear patterns begin to emerge in Table 1. Combinations with "to just," "to really," and "to actually" are much more common in spoken registers than they are in written registers. Within registers, they are much more common in the nonacademic ones. Similarly, "to better," "to fully," and "to effectively" are more common in written registers than they are in spoken ones, and much more common in academic registers. This suggests that lexical patterns or chunks involving the split infinitive are register specific and seem to form a continua of decreasing frequencies from spoken to academic for the first three, and a continua of increasing frequencies for the last three.

The results for the first three combinations in terms of register distribution are not particularly surprising; the adverbs *just*, *really*, and *actually* are much more common in spoken and nonacademic written registers than they are in academic registers, as shown in Table 2.

However, Table 3 shows that the frequencies of "to really," "to effectively," and "to better" seem to be more than a function of the distributions of frequencies of the

Table 1. "*to* + adverb" Combinations in the Corpus of Contemporary American English

to + adverb combination	Spoken	Fiction	Magazine	Newspaper	Academic
to just	32.1	8.3	7.5	8.7	1.4
to really	37.7	3.6	7.5	7.7	2.3
to actually	20.9	3.4	4.6	4.0	2.8
to better	1.8	1.1	3.8	3.7	12.6
to fully	2.5	1.3	4.1	3.6	6.7
to effectively	0.6	0.1	1.3	1.1	6.2

Table 2. Comparison of the Normalized Frequencies of Adverbs in the Six "*to* + adverb" Combinations

Adverb	Spoken	Fiction	Magazine	Newspaper	Academic
just	3,217.7	2,168.9	1,430.8	1,344.2	464.7
really	1,656.8	586.1	435.4	435.8	133.9
actually	617.5	187.1	206.1	133.6	145.5
better	418.5	423.5	471.8	419.7	298.9
fully	41.5	42.2	71.7	50.0	104.3
effectively	20.6	5.2	33.9	22.5	84.4

adverbs alone. The ratios in this table result from dividing the number of times that the adverb alone occurs more often in spoken versus academic registers (or vice versa) between the number of times that the "*to* + adverb" combination occurs more often in spoken versus academic registers (or vice versa). As noted in the method section, ratios closer to 1 indicate that the distribution of the "*to* + adverb" pattern varies in expected ways along with the variation in adverb frequency, whereas ratios higher than 1 indicate that the pattern occurs more frequently than would be expected based on the adverb frequency alone.

As seen in Table 3, "to just" seems to be much more common in spoken registers than the distribution of "just" would suggest. "To better" shows a much stronger register association not only because its ratio is much higher but also because its frequency runs counter to expectations. "Better" is much more common in spoken and nonacademic registers than it is in academic registers. Based on frequencies alone, it could be expected that "to better" would occur more often in spoken registers, but it does not. The high number of occurrences of "to better" thus speaks of a strong register-motivated association between the two words.

As noted in the method section, verb collocates of "to better" were searched first by MI and then by raw frequencies. Table 4 shows the MI results, although frequencies are also provided. The second to last column contains pairs of numbers showing (1) the results for avoidance of the pattern that were obtained by searching for the verb within a nine-word window to the right and a one-word window to the left and (2) the

Table 3. Frequency Ratios

Adverb/split pattern	Ratio
just/to just	3.8
really/to really	1.32
actually/to actually	1.75
better/to better	5—inverse direction
fully/to fully	1.06
effectively/to effectively	2.52

Table 4. Verb Collocates of "to better" Arranged by Mutual Information

Verb	Mutual information	Overall frequency	Frequency of avoidance	Use/avoidance ratio
align	9.80	16	0	—
understand	8.88	642	285	2.25*
coordinate	8.55	16	10/3	5.33
comprehend	8.52	13	6/4	3.25
inform	8.50	25	10/5	5.0
integrate	8.19	16	3/2	8.0
prepare	8.16	65	24/8	8.12
educate	8.02	18	16/5	3.6
accommodate	7.67	19	14/3	6.3
manage	7.67	42	50/20	2.1
assess	7.51	28	15/2	14.0

*Likely to be skewed in favor of avoidance, see explanation below.

results for avoidance after the clauses from (1) were analyzed by hand to determine whether they really were instances of avoidance. For example, for "coordinate," the numbers 10/3 show (1) and (2), respectively. The last column on the right, the use/avoidance ratio, shows how many more times the use of a pattern is more common than its avoidance. A higher use/avoidance ratio suggests that the pattern is more accepted by speakers as it is less likely that it was avoided in the data. Note that the number used to calculate this ratio was the second number (2) in the pair in the previous column when such a number existed. Also note that (2) was not obtained for "understand" because "understand" occurs more than fifty times. As a result, the use/avoidance ratio for "to better understand" is likely to be strongly skewed in favor of the avoidance of the pattern. This likelihood is shown with an asterisk (*) placed next to the use/avoidance ration in Table 4 and all subsequent tables.

Table 4 shows that some of the verbs having high MIs also have high use/avoidance ratios. These include patterns such as "to better align" (which does not really have a ratio because it is never avoided in the data), "to better assess," "to better prepare," "to better integrate," and "to better accommodate." The most frequent pattern,

Table 5. Verb Collocates of "to better" Arranged by Frequency

Verb	Raw frequency	Frequency of avoidance	Use/avoidance ratio
understand	641	285	2.25
serve	75	66	1.13*
prepare	65	24/8	8.12
meet	51	38/1	51.0
protect	45	51	0.88*
manage	42	49/25	1.68
reflect	31	12/4	7.75
assess	28	15/2	14.0
define	26	8/2	13.0
inform	25	10/5	5.0

*Likely to be skewed in favor of avoidance, see explanation below.

"to better understand," which has the second highest MI, is also the one with the low-est use/avoidance ratio, but, as explained, a detailed analysis of each occurrence would very likely result in fewer instances of avoidance.

Table 5 above shows the results of the frequency-focused search. Some patterns are found here that do not emerge in the MI-focused search, such as "to better serve," "to better meet," "to better protect," "to better manage," "to better reflect," and "to better define." The use/avoidance ratio for "to better serve" and "to better protect" is likely to be skewed in favor of the avoidance because of a lack of detailed analysis. The most significant patterns that emerge are "to better meet," which is only avoided once, and "to better define" which has a very high use/avoidance ratio.

Clearly, "to better understand" is the most frequent pattern, although not the one with the highest MI or the highest use/avoidance ratio. Nevertheless, the high MI suggests a very strong lexical bond between "to better" and "understand." As shown in Table 6, "understand" is the most common verb collocate of "to better" in all registers, and "to better understand" is particularly strong in academic registers.

As seen in Table 6, "to better understand" is the most frequent "*to better* + verb" combination in all the COCA registers. The association of "to better" with "under-stand" in academic registers is particularly striking in light of the fact that "understand" is much more common in other registers (Table 7). In other words, the association forming this pattern cannot be a product of a higher frequency of occurrence of "under-stand" in academic registers, and it is thus evidence of strong bonding, as shown by the high MI score.

Results for "to not"

According to Curme (1931), the proscription against the split infinitive applies strongly when "not" is the particle that splits the infinitive. However, data from COCA and the Time Magazine Corpus suggest that "to not" has become more frequent in the past decade, even if it is still far less common than "not to." Table 8 shows the

Table 6. Frequent Verb Collocates of "to better" across Registers

Register	Verb	Raw count
Spoken	understand	47
	protect	11
	inform	4
	prepare	4
	serve	3
Fiction	understand	11
	hear	9
	see	9
	serve	4
	accommodate	3
Newspaper	understand	56
	serve	18
	prepare	17
	manage	16
	protect	13
Magazine	understand	117
	serve	14
	prepare	11
	manage	9
	protect	7
Academic	understand	410
	meet	42
	serve	36
	prepare	32
	reflect	19

Table 7. Normalized Frequencies of "understand" across Registers

Register	Normalized frequency of "understand"
Spoken	324.7
Fiction	247.5
Magazine	147.0
Newspaper	126.1
Academic	199.5

normalized frequencies of "to not" versus those of "not to" in the different registers of the COCA.

The numbers reflect how many times the structures appear per every million words in this corpus containing more than 300 million words. Clearly, "to not" is more common in the spoken language (which in the COCA consists primarily of news programs) than it is in any other register. At the same time, "not to" is several times more

Table 8. Comparison of Frequencies of "to not" and "not to"

Spoken		Fiction		Magazine		Newspaper		Academic	
To not	Not to	To not	Not to	To not	Not to	To not	Not to	To not	Not to
23.9	179.6	5.7	234.5	7.0	160.9	10.5	158.9	6.4	143.9

Table 9. Right-Hand Verb Collocates of "to not" Resulting from Mutual Information–Focused Search

Verb	Raw frequency	Mutual information	Avoidance	Use/avoidance ratio
alienate	5	9.47	18/16	0.31
offend	6	8.81	92	0.06
interfere	6	7.41	124	0.04
cooperate	6	7.20	44	0.13
allow	47	6.82	197	0.23
worry	28	6.64	909	0.03

common than "to not" in all registers. In particular, as Tables 8 and 9 show, "to not" tends to co-occur with certain verbs to the left.

The most significant of these right-hand collocates in terms of MI is "alienate," as in "to not alienate." It is also significant in that it has the highest use/avoidance ratio in a group of items where none is used in the split form more than it is with the nonsplit form, as in "not to worry." In terms of frequency, "to not allow" is the most frequent items and also has a relatively high use/avoidance rate.

Table 10 shows the right-hand collocates, that is, "to not do." "Allow" once again has the highest use/avoidance ratio. The rest of the verbs are all very common ones and are used much more frequently with "not to" than they are with "to not."

Table 11 combines the results of an MI-focused search and a frequency-focused search for left-hand collocates of "to not," including items other than verbs. These results are combined in a single table and organized by frequency because of the fact that the lemmas of a single verb had to be grouped to avoid repetition. The item showing the least avoidance is "committed," followed by "seem" and "tend." The highest frequency, however, corresponds to "try" and the highest MI is for "tendency," closely followed by the related word "tend."

Table 12 shows the diachronic variation in the frequency of "not to." A modest but steady increase in frequency of occurrence is apparent, suggesting a trend toward greater use or less editorial intervention on this front.

Discussion and Conclusion

The first part of this article has shown that there seem to be multiple written sources for the proscription of the split infinitive. One of them, "P," embodies a familiar ideology

Table 10. Left-Hand Verb Collocates of "to know" Resulting from Frequency-Focused Search

Verb	Raw frequency	Avoidance	Use/avoidance ratio
be	396	7,253	0.05
have	248	1,918	0.12
let	113	961	0.11
do	92	1,504	0.06
get	90	1,051	0.08
go	61	1,019	0.05
give	53	509	0.1
make	50	907	0.05
allow	47	147	0.23
know	43	410	0.1

Table 11. Right-Hand Collocates of "to not"

Right-hand collocate	Raw frequency	Mutual information	Avoidance	Use/avoidance ratio
try	77	4.21	1,632	0.04
tend	36	6.03	430	0.08
admit	24	5.56	0	—
choose	24	5.12	594	0.04
careful	22	5.90	1,653	0.01
decide	22	4.18	1,956	0.01
ought	18	5.32	617	0.02
committed	17	5.41	9	1.8
seem	17	3.72	172	0.09
tendency	12	6.13	531	0.02

Table 12. Normalized Frequency of "to not" across Time in the Corpus of Contemporary American English

1990–1994	1995–1999	2000–2004	2004–2008
8.4	10.2	11.3	13.8

positing that the nonstandardness of a feature derives from, and is a reflection of, the lower education, lower intelligence, or base character of its users. Importantly, another source, Robert Taylor, links the proscription to an ideal of Teutonic purity by explicitly comparing English to German and by implying that the split infinitive makes English depart from German models. As noted previously by Bailey, there is

no support for the Latin origin hypothesis in the written records, but there is some support for a German origin hypothesis.

The second part of the article has shown that the advice offered by contemporary usage guides can be contradictory across sources and even within the same entry in the same source. Importantly, such advice is for the most part not founded on in-depth explorations of usage. Rather, it is predicated on very subjective ideas of clar-ity and elegance, with little by way of definition of those. Exceptions to this are Crystal's suggestion that the split infinitive is acceptable when it preserves what he proposes is the natural rhythm of English, akin to the rhythm of iambic verse, and Quirk et al.'s remarks about the high frequency of subjuncts as the splitting element in split infinitives.

The third part of this article, the corpus-based study, has attempted to characterize some lexical patterns involving the split infinitive that might be emerging at present as well as their variation across registers. Some clear patterns, or chunks, have been observed for the split infinitive when adverbs are the splitting element. Importantly, most of the emerging patterns show clear register preferences: "to just," "to really," and "to actually" clearly prefer the spoken and the less formal (nonacademic) written registers. On the other hand, "to effectively" and "to better" show strong associations with academic registers, with "to fully" showing a weaker one. The clearest three-word pattern in terms of frequency, strength of semantic bonding, higher than random occurrence and relatively high use/avoidance ratio is "to better understand." Beside "to better understand," other more or less clear patterns emerging in this study include "to better align," "to better meet," "to better assess," and "to better prepare." Importantly, a single criterion alone, either MI or frequency, would have left out some of them, which speaks to the importance of an approach that combines both MI and frequency to characterize word or phrase collocates.

"To better understand" is strongly associated with academic registers. In general, patterns involving the split infinitive are clearly differentiated across registers, which runs counter to some expectation that perhaps patterns originating in oral language, which is less likely to follow prescriptivism, would somehow spill over to written registers. Instead, the written register perceived by many to be the most rigid, the aca-demic register, has given rise to its own split infinitive forms, namely "to better" and its verb collocates, as well as "to effectively," whose verbs collocates were not addressed in this study. Thus, it does not seem to be the case that the more lax mores of oral language influence the written language when it comes to the split infinitive.

Nor does it seem to be always the case that, as Crystal suggests, the split infinitive is a matter of preserving the natural strong–weak–strong–weak beat of English sylla-bles. Crystal's pattern does apply to "to *bet*ter un*der*stand" and "to *bet*ter *meet*." However, the pattern doesn't apply to "to *bet*ter prep*are*," "to *bet*ter as*sess*," or "to *bet*ter al*ign*." For those, the rhythm is rather weak–strong–weak–weak–strong. By contrast, Quirk et al.'s assertion that the split infinitive is most common with subjuncts holds in the results of this study, as the six most common splitting adverbs are sub-juncts that perform a grading or focusing function.

The differences in the distribution of patterns across registers suggest that another principle, that of idiomacity, might be at work in the preferences for certain forms to be split. John Sinclair, one of the first proponents of the principle of idiomacity in English, observes that language is largely systematic in that "a language user has available to him or her a large number of semi-preconstructed phrases that constitute a single choice, even though they might appear to be analyzable into segments" (Sinclair 1966:110). Some kinds of preconstructed phrases have been shown to associate strongly with specific registers (Biber, Conrad, & Cortes 2004). I suggest that the split infinitive forms above are instances of a similar phenomenon. In particular, "to better understand" may have emerged as a collocational pattern in academic discourse in connection with the purported mission of academia, namely to achieve better understandings of phenomena or situations. Given the centrality of better understandings, it makes discursive sense to keep the words together. In addition, "to better understand" is usually followed by longer nominal clauses or postmodified noun phrases. Placing the adverb after the noun phrase would stretch discourse processing capacities, which can be avoided by keeping the adverb close to the verb. In addition, the pattern "to *under*stand *be*tter" would violate Crystal's principle of euphony, which may explain its avoidance.

The situation for "to not" is, in some ways, clearer: it almost never is the preferred form. When it is, as in "admit to not doing something," the particle "to" is functioning as a preposition in a prepositional verb than as a marker of the infinitive mood. But, even then, a potential rule that such prepositional verbs always "split" their infinitives is precluded by the existence of sentences such as "he is committed not to change the tax policy at all." In other words, the characterization of what a split infinitive is with "to not" is problematic. Some patterns that seem to be emerging are "to not allow" and "tend/a tendency to not." But for those and all other collocates of "to not," the use of "not to" is always much higher. This suggests that, while "to not" may prefer such words, the words themselves do not prefer "to not." Rather, they prefer "not to." Nevertheless, "to not" seems to be slowly becoming more frequent, which suggests that we might see collocational patterns for "to not" more frequently in the future.

Declaration of Conflicting Interests

Funding

Notes

1. Quirk et al. (1985:566) categorize adverbials in four categories: adjuncts, subjuncts, disjuncts, and conjuncts. Subjuncts are adverbials "which have, to a greater or lesser degree, a

subordinate role in comparison with other clause elements." Examples of subjuncts include "badly," "well," and "rather."

2. Cutoff points defining what counts as a lexical bundle have been set at different frequencies by different studies.

Corpora

Davies, Mark. Corpus of Contemporary American English. Brigham Young University. http://www.americancorpus.org.

Davies, Mark. Time Magazine Corpus. Brigham Young University. http://corpus.byu.edu/time/x.asp.

References

Alford, Henry. 1863. *A plea for the queen's English*. London: G. Bell & Sons.

Altenberg, Bengt. 1990. Speech as linear composition. In G. Caie, K. Haastrup, L. Jakobsen, J. Neilsen, J. Sevaldsen, H. Spect, & A. Zettersten (eds.), *Proceedings from the fourth Nordic conference for English studies*, vol. 1., 133-143. Copenhagen: University of Copenhagen.

American heritage dictionary of the English language, 4th edn. 2000. Split infinitive, def. 59. New York: Houghton Mifflin,

Bailey, Richard. 2006. Talking about split infinitives. *Michigan Today News-E*, June 2006, http://www.ns.umich.edu/MT/NewsE/06_06/words.html (11 August 2010).

Biber, Doug & Federica Barbieri. 2007. Lexical bundles in university spoken and written registers. *English for Specific Purposes* 26. 263-286.

Biber, Doug, Susan Conrad, & Viviana Cortes. 2004. If you look at. . . . Lexical bundles in university teaching and textbooks. *Applied Linguistics* 25. 371-405.

Crystal, David. 2006. *The fight for English: How language pundits ate, shot, and left*. Oxford, UK: Oxford University Press.

Curme, George O. 1931. *Syntax*. Boston: D. C. Heath.

Fowler, Henry W. & Robert Burchfield. 1998. *The new Fowler's modern English usage*. Oxford, UK: Clarendon.

Hall, Fitzgerald. 1882. On the separation, by a word or words, of "to" and the infinitive mood. *American Journal of Philology* 3. 17-24.

Howard, Godfrey. 1997. *The Macmillan good English handbook*. London: Macmillan.

Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge, UK: Cambridge University Press.

Lowth, Robert. 1762/1967. *A short introduction to English grammar*. Menston, UK: Scholar Press.

Milroy, James & Lesley Milroy. 1991. A*uthority in language: Investigating language prescription and standardisation*. 2nd edn. London: Routledge.

Murray, Lindley. 1792/1968. *English Grammar*. Menston, UK: Scholar Press.

Myers, Marshall. 2002. A new look at infinitives in technical communication. *Technical Writing and Communication* 32. 59-65.

New York Public Library Writers' Guide to Style and Usage. 1994. New York: Collins.

Oxford English Dictionary, 2nd edn. 1989. Infinitive, def. B1. Oxford, UK: Oxford University Press.

P, letter to the editor. 1834. Inaccuracies of diction. Grammar. Cornell University Library. http://digital.library.cornell.edu/cgi/t/text/text-idx?c=nwen;idno=nwen0007-6 (15 October 2008).

Partridge, Eric. 1999. *English usage and abusage*. 1st American edn. New York: Norton.

Peters, Pam. 2004. *The Cambridge guide to English usage*. Cambridge, UK: Cambridge University Press.

Pinker, Steven. 2009. Oaf of office. *New York Times*, January 21. http://www.nytimes.com/2009/01/22/opinion/22pinker.html?_r=1 (23 June 2010).

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.

Sinclair, John. 1966. Beginning the study of lexis. In Charles Ernest Bazell, J. C. Catford, M. A. K. Halliday, & Robert H. Robins (eds.), *In memory of J.R. Firth*, 410-430. London: Longman.

Sorenson, Sharon. 1997. *Webster's new world student writing handbook*. 3rd edn. New York: MacMillan.

Strunk, William & Elwyn B. White. 1979. *The elements of style*. 3rd edn. New York: MacMillan.

Taylor, Robert. 1840. Preface and additional notes. In John Tooke's *The diversions of Purley*, v-liv. London: Robert Taylor.

Visser, Fredericus. 1984. *An historical syntax of the English language*. Leiden, Netherlands: E. J. Brill.

Webster's Dictionary of English usage. 1989. Split infinitive. Springfield, MA: Merriam-Webster.

Williams, Joseph. 2007. *Style: Lessons in clarity and grace*. 9th edn. New York: Pearson Longman.

Bio

Moisés D. Perales-Escudero is a doctoral candidate in the Joint Program in English and Education at the University of Michigan. His research focuses on discourse analysis and its pedagogical applications.