

The English Split Infinitive across Genres

Supakorn Phoocharoensil
Thammasat University

Abstract

The English split infinitive has been prominent among linguistic topics of scholarly debate for more than a century. While prescriptivists tend not to accept this grammatical structure, labeling it as a non-standard form, modern L1-English speakers often ignore this old-fashioned Latin-based rule. This corpus-based study investigated split infinitives occurring in the eight different genres of the Corpus of Contemporary American English (COCA), focusing on single-adverb splitters and categorizing split infinitives according to the functions they serve. The results indicate that infinitive splitting is the most common in two new related web-based genres of COCA, i.e. blogs and webpages, which implies the high degree of informality associated with both text types. It was also discovered that *to better understand* was the most frequent in blogs, webpages, and academic texts. The principal function of split infinitives in blogs and webpages is to mark a relationship with other possible circumstances, while those of academic texts and spoken language are to modify a gradable verb and to mark completion respectively.

Key words: the English split infinitive, COCA, American English, blogs and webpages, distribution across genres

1. Introduction

As English is a language with a very long history, grammar rules seem to be always changing. One of the English grammatical constructions that has been a controversial issue for grammarians and users for over a century is the split infinitive, e.g. *to boldly go* or *to fully understand* (Mitrascu, 2009). Prescriptive linguists who formulate grammar rules and enforce the correct usage of those rules undoubtedly reject the phenomenon whereby an infinitive structure comprises an element, e.g. an adverb, between *to* and a verb in its base form. Descriptive linguists, who by contrast objectively analyze and describe how a language like English is actually used by a speech community, mostly allow for infinitive splitting, i.e. a linguistic construction that quite a few native speakers of English show preference for in certain situations, e.g. movie blurbs in which natural rhythms of English are the norm (Crystal, 2003), or in ambiguity reduction in writing (Calle-Martin and Miranda-Garcia, 2009; Schwarz & Smitterberg, 2020).

A number of corpus-based studies on split infinitives have reported on a steady increase of this structure in native speakers of

English, e.g. British English (Mikulova, 2011; Mitrascu, 2009) or American English (Albakri, 2005; Johansson, 2015; Mikulova, 2011; Mitrascu). Some studies focused on split infinitives in non-native varieties of English (e.g. Calle-Martin & Romero-Barranco, 2014). Although the split infinitive is assumed to be characteristic of spoken English (Carter & McCarthy, 2006), the construction has been continuously gaining more popularity in academic genres (Johansson, 2015). Evidently, the split infinitive structures in these two distinct text types behave differently. The research gap that the current study also aimed to bridge concerns the occurrences of split infinitives in web-based English, and since the latest version of the Corpus of Contemporary American English (COCA) includes three more genres, two of which, namely 'blogs' and 'webpages', represent English as it is widely used on the Internet, this study of American-English split infinitives relied on data from the updated version of COCA (Davies, 2020).

The next section deals with the definition of split infinitives, reasons for and against the use of this construction, and findings of past studies relevant to split-

infinitive occurrences in native speaker English.

2. Review of Literature

2.1 Defining the English Split Infinitive

The split infinitive in English consists of the infinitive marker *to* and a splitter, which is normally a single adverb, such as *really* in (1), or the negative particle *not*, as in (2), followed by a verb in its base form (Huddleston and Pullum, 2002).

(1) She seems *to really like* it.
(Oxford advanced learner's dictionary, 2015, p. 1507)

(2) They will give young people yet another reason *to not go* to church.
(Mikulova, 2011, p. 28)

While the split-infinitive construction is usually believed to be characteristic of modern English, it has been discovered that, historically speaking, the structure originated in Middle English, around the 13th century (Calle-Martin, 2015). More than sixty famous authors of literature, science, and political discourse publishing in that period produced different forms of the split infinitive (Perales-Escudero, 2010). Despite the fact that the popularity of split infinitive use declined during the 16th and 17th centuries, it became more common in the nineteenth and twentieth centuries (Mikulova, 2011). There have been conflicting linguistic views on acceptance of the split infinitive, which will be discussed below.

2.2 To Split or Not to Split an Infinitive: Criticism and Support

On one hand, prescriptive linguists who propose and strictly enforce a set of rules on language based on how they think language should be used, are strongly critical of and solidly against the split infinitive in English, labeling it as a forbidden, ungrammatical structure. For these grammarians, there are reasons, mostly unjustified, to prohibit infinitive splitting (Mitrasca, 2009). They are of the opinion that the infinitive is a fixed language unit that must not be broken or split with any inserted element. Prescriptivists,

moreover, ban the split infinitive by devising a rule associated with Latin, a once dominant and prestigious language. The heavily-Latin-based English grammar disallows the split infinitive simply because such a construction does not exist in Latin. In other words, grammarians under the influence of Latin grammar arbitrarily proscribe splitting an infinitive in English due to the non-existence of such a construction in Latin (Peters, 2006).

Descriptive grammarians, on the other hand, have made several objections to the prescriptive old-fashioned rule that forbids infinitive splitting. Descriptive grammar, unlike prescriptive grammar, has its goal of describing the usage of a language by native speakers rather than control their usage as prescriptive grammar has an attempt to. Among descriptive linguists, a structure like *to go* is not considered to be an inseparable linguistic unit; the real infinitive is *go*, whereas *to* only functions as a linking particle (Trask, 2007). Consequently, it should be possible for *to go* to be parted by an element, e.g. the adverb *boldly*, as in *To boldly go where no man has gone* before, a well-known line from *Star Trek*. Not only is this particular phrase *to boldly go...* grammatically acceptable, but also it follows the natural rhythm of English and results in a richer flavor than the banal non-split counterpart, i.e. *to go boldly* (Crystal, 2003). Crystal's viewpoint is in line with Strumpf and Douglas (2004), who allow infinitive splitting on some occasions despite the fact that it is potentially regarded as a mistake in some grammarians' opinion, claiming that split infinitives are sometimes permitted and employed by well-known writers to place an emphasis on certain points or create poetic power.

The rule proscribing the split infinitive in English is primarily derived from a false analogy with Latin grammar. Put simply, many rules of the English grammar have been formed on the basis of Latin. As a matter of fact, living up to the standards of Latin is linguistically inappropriate since the two languages are originally different in a number of ways. Regarding the infinitive structure, as Huddleston and Pullum (2002) remarked, the infinitive phrase *to love* in English is translated into only a single indivisible word *amare* in Latin. This accounts for why in Latin it is

unlikely that a splitter such as an adverb could occur within the infinitive. By contrast, to separate the infinitive marker *to* and the following base-form verb is easier or more possible. According to Mitrasca (2009), in comparison to English verb phrases in which an adverb can be positioned between an auxiliary and a main verb as in (3) where the adverb *actually* occurs between the auxiliary *did not* and the main verb *talk*, a construction such as *to seriously consider*, in which the adverb splitter seriously comes between *to* and the main verb, should also be viewed as grammatical.

(3) We *did not actually talk* to him.
(Mitrasca, 2009, p. 105)

It is also important to use the split infinitive for clarity of meaning. According to Calle-Martin and Miranda-Garcia (2009), splitting an infinitive contributes to reduction of vagueness and ambiguity. The different sentence positions of the adverb *really* in (4) and (5) cause a difference in meanings. The meaning of (4) is ‘it is important that you watch him, while that of (5) is ‘you need to watch him very closely’. Strictly complying with the prescriptive rule which disallows splitting an infinitive would only allow the speaker to convey meaning through (4) although the actual meaning being conveyed is (5). In such as case, using a split infinitive is necessary to preserve the exact intended meaning.

(4) You *really* have to watch him.
(5) You have to *really* watch him.
(<https://en.oxforddictionaries.com/grammar/split-infinitives>)

Due to the aforementioned reasons, the split infinitive in English should no longer be treated as a grammatical error or linguistically ill-formed structure, given that the arguments in support of banning split infinitives are relatively weak. All things considered, there are no convincing grammar rules requiring that the infinitive marker *to* always stay adjacent to the verb that follows.

The next section discusses the occurrences of split infinitives in spoken and

written modern English as shown in corpus-based data.

2.3 The Split Infinitive and Native-Speaker English: Evidence from Language Corpora

With language corpora representing English native speakers’ authentic use of their mother tongue, researchers can now access linguistic evidence through real-English data for their own studies. There is no longer a need to depend on individual native speakers’ intuition, which often turns out to be inaccurate or biased, to determine the degree a language construction is used (Cheng, 2012). This way, the occurrence of split infinitives in native speakers’ speech and writing can be examined with more precision by means of corpus-informed data.

Albakry (2005)’s comparative study between two American newspapers showed that *USA Today* allows split infinitives to a larger degree than *New York Times* since 10% of the split-infinitive construction were found in the former, compared to only 6% in the latter. Additionally, adverbs were shown to be the most common splitters in both newspapers. More specifically, the typical adverbs functioning as splitters were adverbs of manners, such as *significantly*. The main purpose for infinitive splitting is meaning clarification as well as emphasis.

One of the most well-cited English grammar references, i.e. Cambridge Grammar of English, was compiled by two corpus linguists Ronald Carter and Michael McCarthy (2006), who collected data from the Cambridge International Corpus (CIC), representative of British English, American English, and other major Englishes. Carter and McCarthy maintained that split infinitives, e.g. *to actually not like* and *to automatically laugh* in (6), although often considered wrong, showing bad style, or even inappropriate in written English, are commonly used in spoken English.

(6) It’s very common *to actually not like* the Birmingham accent, isn’t it? People tend *to automatically laugh* at it.
(Carter & McCarthy, 2006, p. 596)

A more corpus-based study exploring split infinitives in British and American

English was conducted by Mitrasca (2009). In this study, British English was represented by data from the British National Corpus (BNC) and data on American English was drawn from the Corpus of Contemporary American English (COCA). A significant difference between British and American usages of split infinitives was clearly demonstrated. That is, the split infinitive in American English occurred approximately three to four times more than in British English, which implies more popularity of this linguistic construction among Americans. Double adverbs, e.g. *at least, not only, sort of, kind of*, are commonly used as splitters in American English, whereas they were rare in the British counterpart. Compound-adverb splitters are notably prevalent in American English. More precisely, *more* is the most common adverb to constitute compound splitters, e.g. *to more fully understand, to more accurately reflect*, etc. On the whole, Mitrasca pinpoints the obsolescence of the prescriptive rule forbidding split infinitives as corpus-informed data clearly confirm the split infinitive prevalence in mainstream varieties of English.

A distinction between how the British and Americans use split infinitives has been made clearer by Mikulova (2011), whose study was also based on data from the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). In support of Mitrasca (2009), Mikulova reported a higher frequency of split infinitives in American English than in British English. With a closer investigation into both corpora, split infinitives in BNC spoken data outnumber those in the written one. However, the COCA data indicate a nearly equal distribution of split infinitives in both spoken and written American English. The study also reveals the most frequent adverb splitters in the two corpora. In particular, the top-ten adverb splitters in BNC and COCA, ranked in frequency, are as follows (Mikulova, 2011, p. 27):

Top-ten adverb splitters in BNC: *actually, just, really, even, further, fully, completely, always, finally, and better*

Top-ten adverb splitters in COCA: *just, really, actually, better, even, further, fully, kind of, always, and simply*

Apart from adverb splitters, the negative particle *not* is also used as a splitter with less frequency than the adverb counterpart, as in “...you soon learn how to not make the same mistakes again.” (Mikulova, 2011, p. 28). It is worth noting that *not* placed between the infinitive marker *to* and the verb sounds more negative than when it appears immediately before *to* (Fitzmaurice, 2000). Complex adverbials also act as splitters, such as *just sort of, sort of just, all of a sudden*, etc. These complex-adverbial splitters occur in American English as opposed to British English. In addition, there is a preponderance of comparative adverbials using *more*, e.g. *to more fully* or *to more closely*, in the American corpus data over the British one.

Johansson (2015) also investigated the occurrences of split infinitives in American English based on the data from the Corpus of Historical American English (COHA) and the Corpus of Contemporary American English (COCA). COHA was consulted for the historical perspective of split infinitives, whereas COCA informed the present-day usage of this construction. It was discovered that the split-infinitive frequency increased rapidly over periods of time. In particular, the split infinitive appeared with the highest frequency in informal spoken English. More surprisingly, it was academic English, i.e. the genre most associated with a high degree of formality, in which this infinitive structure saw the largest rise in prevalence, with *to better understand* being the most frequent in this genre. The continuously rising frequency of split infinitives, which is in line with other studies, demonstrate the widespread use of split infinitives across all genres and different levels of formality, confirming “...the split infinitive is becoming more and more a standard feature for every new generation of English speakers” (p. 22). The higher degree of acceptance of split infinitives in academic writing is attributed to the facilitation of semantic clarity and transparency by positioning the adverb right before the verb for ambiguity avoidance. As a conclusion, due to

its higher level of popularity and frequency, teachers should no longer consider prohibiting students' use of this infinitive structure.

The present study aimed at exploring the split infinitives in American English with emphasis on the distribution of this grammatical construction across a variety of genres present in the Corpus of Contemporary American English (COCA). The study examined only the split infinitives made up of a single-adverb splitter, focusing on 1) the genres with the highest frequency of this construction and 2) a comparison between academic and spoken English, i.e. the two text types in which the most notable differences of the split infinitive use are expected.

3. Methodology

COCA was consulted for all frequencies of the split infinitives in American English in eight different genres, namely, five traditional genres, i.e. spoken, fiction, popular magazine, newspaper, and academic texts, and three new genres available as of March 2020, i.e. TV and Movie subtitles, blogs, and webpages. At the time in which data collection took place, COCA was composed of over one billion words derived from 485,202 texts, including 24-25 million words each year from 1990-2019 (Davies, 2020). To find the total occurrences of single adverb splitters placed between the particle *to* and the infinitive, the search string was *to [r*] [v?i*]*, in which *to* and *[v?i*]* stands for the infinitive marker *to* and a verb in its infinitival form respectively, while *[r*]* extracts a single adverbial.

Among the eight genres in COCA, academic texts represent the most formal

English as data was drawn from almost 100 different peer-reviewed journals, covering a wide range of disciplines. In contrast, the most informal is TV/Movies subtitles and spoken language, with the former being as informal as or more informal than the latter. It is also important to note the two newly added genres, i.e. blogs and webpages, are both taken from the Corpus of Global Web-Based English (GloWbE) and characterize American English on the Internet. While nearly all the texts in 'blogs' are restricted to only blogs, those on 'webpages' are representative of the 'General' texts from the United States in the GloWbE corpus and some of the texts cover blogs since it was impossible to exclude the blog data at the time the corpus was compiled. There could therefore be some similarities of the structure being investigated in the data shown in these two subcorpora.

It is worth noting that, due to the differences in size of the subcorpora, the normalized frequency (i.e. per million) instead of the raw frequency was used for comparison between genres. The latest version of COCA is of an acclaimed quality due to the fact that all the occurrences of the aforementioned query did not include any grammatical construction other than the split infinitive. In the next step, the *to + adverb + verb* occurrences which are in the top-20 frequency list presented in each of the eight genres in COCA were listed. An analysis of the extracted split infinitives representing each particular genre can, to a certain extent, account for the genre-specific use of split infinitives (Johansson, 2015).

4. Results and Discussion

4.1 Distribution of Split Infinitives across COCA Genres

Table 1 Frequency of Split Infinitives in the Eight Genres of COCA

Genre	Frequency	Per Million
Blogs	6,286	50.0891
Webpages	4,821	37.1133
Spoken	1,150	9.0269
Academic	857	7.0833
TV/Movies	230	1.7966
Magazines	225	1.7667
Newspapers	87	0.7075

Fiction	10	0.0836
TOTAL	13,666	

As seen in Table 1, the genres in which split infinitives occur with the highest frequency are ‘blogs’ (50.0891 per million) and ‘webpages’ (37.1133 per million), both of which share a number of similarities due to the fact that their data are representative of web-based English, with ‘blogs’ containing approximately 125 million words and ‘webpages’ around 130 million words. To be more precise, the data of these two new COCA genres were obtained from American English in GloWbE, and some of the texts in ‘webpages’ also inevitably included blogs at the time the corpus was created (Davies, 2020). Before COCA introduced these two new genres in 2020, ‘spoken’ was found to contain the highest tokens of split infinitives, clearly indicating that infinitive splitting is characteristic of spoken American English (Johansson, 2015; Mikulova, 2011). The preponderance of split infinitives in ‘blogs’ and ‘webpages’ over the other genres in COCA shows that informal English is common in writing blogs and webpages, in which informal and reader-friendly language is expected.

The frequency of split infinitives in the ‘spoken’ genre, composed of transcripts of unscripted conversation from more than 150 different TV and radio programs, was ranked third (9.0269 per million) in the current version of COCA. This implies that infinitive splitting is more permissible in spoken, colloquial English than in the text types where high level of formality is the norm, e.g. academic texts, which is in line with previous studies (e.g. Carter & McCarthy, 2006; Johansson, 2015; Mikulova, 2011). The COCA genres also concerned with informal English are ‘TV/Movies’ and ‘magazines’. Both are very close in split infinitive frequency, with ‘TV/Movies’ and ‘magazines’ containing 1.7966 per million and 1.7667 per million, respectively. Like those in ‘spoken’ texts, the texts constituting the corpus of TV and Movie subtitles represent informal English. More specifically, the degree of informality in these TV and movie subtitles is more or less equal to or, in some cases, even higher than that in the ‘spoken’ genre. The

corpus of ‘magazine’, in a similar vein, comprising almost 100 different magazines from various contexts, e.g. health, home and gardening, women, financial, religion, sports, is a genre where less formal language is expected, and this may explain why split infinitives also exist in magazines.

While the split infinitive is associated with spoken English, it is surprising to see the high frequency of this infinitive structure in ‘academic’ genre of COCA (7.0833 per million), following the ‘spoken’ genre. The academic texts in COCA, with approximately 121 million words, were collected from nearly 100 different peer-reviewed journals. The selected topics cover a broad range of academic fields, e.g. philosophy, psychology, religion, world history, education and technology. The high number of split infinitives in academic texts, where the most formal language is to be expected, confirms Johansson (2015), who reported on the upward trend of this infinitive construction in the academic language of Americans as observed from COCA data between 1990 and 2012.

The last two genres with the lowest frequency are ‘newspaper’ and ‘fiction’ respectively, which also confirms Johansson (2015) in that split infinitives occurred the least in fiction, followed by those in newspapers during 1990-2012. In making the corpus of newspapers, ten different US newspapers were selected, with a proportionate mix between various newspaper sections, such as local news, opinion, sports, finance, etc. The fiction corpus consists of short stories and plays from literary magazines, children’s magazines, popular magazines, first chapters of first edition books 1990-present, and movie scripts. Although the language of fiction is supposedly informal, this is the genre containing the lowest number of split infinitive occurrences, which may imply that infinitive splitting is not preferred in this particular text type. In 4.2, the most common adverb+ verb combinations as part of split infinitives in an individual genre will be presented. The main focus will be on the combinations that specifically characterize each genre.

4.2 The Adverb+Verb Combinations in Different Genres

Table 2 High-frequency Split Infinitive Constructions in Blogs and Webpages

Ran k	Blogs	Freque ncy	Webpages	Frequency
1	TO BETTER UNDERSTAND	253	TO BETTER UNDERSTAND	313
2	TO JUST BE	227	TO JUST BE	180
3	TO ACTUALLY DO	195	TO STILL BE	143
4	TO ALWAYS BE	182	TO ALWAYS BE	136
5	TO STILL BE	165	TO REALLY GET	132
6	TO REALLY GET	155	TO JUST GET	122
7	TO ACTUALLY GET	145	TO ACTUALLY BE	121
8	TO JUST GO	145	TO FULLY UNDERSTAND	104
9	TO ACTUALLY BE	143	TO JUST GO	99
10	TO JUST GET	139	TO ACTUALLY DO	98
11	TO EVEN CONSIDER	106	TO ALSO BE	97
12	TO EVEN BE	102	TO FINALLY GET	83
13	TO JUST SAY	100	TO EVER BE	79
14	TO EVEN THINK	99	TO REALLY BE	78
15	TO ALSO BE	94	TO JUST KEEP	77
16	TO JUST DO	93	TO ACTUALLY GET	75
17	TO REALLY UNDERSTAND	93	TO JUST LET	73
18	TO ACTUALLY MAKE	92	TO EVEN THINK	72
19	TO REALLY BE	90	TO EVEN BE	67
20	TO FINALLY GET	89	TO EVEN CONSIDER	65

According to Table 2, a number of adverb+verb combinations are shared by the split infinitives in ‘blogs’ and ‘webpages’. The combinations found only in the top-20 list of ‘blogs’ are *to just say*, *to just do*, *to really understand*, and *to actually make*, while those appearing in the list of ‘webpages’ but not in the other are *to fully understand*, *to ever be*, *to just keep*, and *to just let*. Interestingly, the two

most frequent combinations in both corpora are *to better understand* and *to just be*; the fourth rank, i.e. *to always be*, is also the same in both genres, and *to still be* and *to really get* are in the top-6 lists of the two corpora. Due to the fact that the data of ‘webpages’ also includes blogs, the same adverb+verb combinations existing in these two text types did not come as a surprise.

Table 3 Functions of Adverb Splitters in Blogs and Webpages

Functions	Blogs		Functions	Webpages	
	adverb splitter	frequency		adverb splitter	frequency
1.mark an inclusive or exclusive relationship with other possible	<i>just (704)</i> <i>even (307)</i>	1011	1.mark an inclusive or exclusive relationship	<i>just (551)</i> <i>even (204)</i> <i>ever (79)</i>	834

occurrences			with other possible occurrences		
2.mark completion	<i>actually</i> (575) <i>really</i> (338)	913	2.act as subjuncts of narrow orientation modifying a gradable verb	<i>better</i> (313) <i>still</i> (143) <i>always</i> (136) <i>fully</i> (104) <i>also</i> (97)	793
3. act as subjuncts of narrow orientation modifying a gradable verb	<i>better</i> (253) <i>always</i> (182) <i>still</i> (165) <i>also</i> (94)	694	3.mark completion	<i>actually</i> (294) <i>really</i> (210)	504
4.act with occurrences that can be imagined as taking place sooner or later along a time scale	<i>finally</i> (89)	89	4.act with occurrences that can be imagined as taking place sooner or later along a time scale	<i>finally</i> (83)	83

The adverb splitters found in the top-20 list of the adverb+verb combinations in both genres were then classified according to Close (1987)'s functions. The most common function, expressed by the adverb splitters *just* and *even* in blogs and webpages, with the exception of *ever* in the latter corpus, was to mark an inclusive or exclusive relationship with other possible occurrences, as shown in (7-10).

(7) Hopefully the little one will pull through and this might end up proving **to just be** a horrible sequence of bad luck events. (blogs)

(8) Some people don't have the opportunity **to just get** together. (webpages)

(9) In fact, I will certainly be disinherited if I were **to even consider** making these again and not inviting her and my father for dinner. (blogs)

(10) I have never had the time **to even try and think** of a way to test these ideas. (webpages)

The findings in Table 3 reveals that the least frequent function of adverb splitters

in the two genres was to act with occurrences that can be imagined as taking place sooner or later along a time scale; the only adverb splitter present in both corpora was *finally*, as in (11)-(12). It is also interesting to note that both blogs and webpages share the same adverb splitters *actually* and *really* to mark completion; moreover, the adverbs *better*, *always*, *still*, and *also* were found to split the infinitive structures in both genres to act as subjuncts of narrow orientation modifying a gradable verb, with *fully*, as exemplified in (13), being among those in the top-20 list of webpages but not the other genre.

(11) What will it take for someone in charge **to finally get off** their high horses. (blogs)

(12) I'm rather glad he seems **to finally be** firmly out of his classical closet. (webpages)

(13) There is no way for them **to fully understand** mighty China market like domestic companies. (blogs)

Table 4 High-frequency Split Infinitive Constructions in Academic Texts, Spoken English, and TV/Movies

Ran	Academic Texts	Frequ	Spoken English	Freq	TV/Movies	Fre
-----	----------------	-------	----------------	------	-----------	-----

k		ency		uency		que
1	TO BETTER UNDERSTAND	791	TO JUST BE	222	TO FINALLY MEET	152
2	TO FULLY UNDERSTAND	29	TO REALLY GET	210	TO BOLDLY GO	78
3	TO FURTHER EXPLORE	10	TO REALLY BE	135		
4	TO SUCCESSFULLY NAVIGATE	9	TO JUST SAY	107		
5	TO STRONGLY AGREE	9	TO ACTUALLY GET	82		
6	TO ACTIVELY ENGAGE	9	TO JUST GET	71		
7			TO REALLY MAKE	52		
8			TO ACTUALLY DO	42		
9			TO ACTUALLY HAVE	41		
10			TO JUST GO	39		
11			TO STILL BE	28		
12			TO JUST DO	21		
13			TO ACTUALLY BE	20		
14			TO REALLY UNDERSTAND	19		
15			TO REALLY HAVE	12		
16			TO BETTER UNDERSTAND	11		
17			TO JUST SIT	10		
18			TO JUST KEEP	10		
19			TO REALLY TAKE	9		
20			TO EVEN BE	9		

As one of the main objectives of this study is to make a comparison between the split infinitives in academic and spoken American English, the most frequent combinations of *to* + *adverb* + *verb* in both genres, with ‘spoken’ English including TV and movies subtitle language, have been listed. While the first twenty most common adverb+verb combinations were extracted from the ‘spoken’ genre, only six and two appeared in the ‘academic’ and ‘TV/Movies’ genres respectively. It is important to note here that although the two split infinitive constructions occurring in the corpus of TV and movies subtitles *to finally meet* and *to boldly go* are characteristic of informal

English, both combinations exist only in TV and movies; the famous combination *to boldly go* is extracted from the introduction to Star Trek *To boldly go where no man has gone before*, which first aired in 1966.

(14) It's so great **to finally meet** you, Danny, in the flesh. (TV/Movies)

(15) Its continuing mission, to explore strange new worlds to seek out new life and new civilisations **to boldly go** where no one has gone before. (TV/Movies)

In the academic texts, only six combinations were discovered, the most frequent of which (791 tokens), i.e. *to better*

understand, as in (16), also occurred in spoken English with far lower frequency (11 tokens), as in (17). It is also noteworthy that this particular most frequent split-infinitive chunk is the most common in blogs and webpages as well. It is apparent that all the adverb splitters except *better* are associated with formal English, e.g. *further*, *successfully*, as shown in (18), and did not exist in spoken English, where informal adverbs prevailed, e.g. *just*, *really*, *actually*, as exemplified in (19).

(16) The goal of this study is **to better understand** the current state of online health resources provided on public library websites. (Academic)

(17) Like, what kinds of people did you seek out in order **to better understand** what day-to-day life was like there? (Spoken)

(18) Additional research is needed **to further explore** the acceptability of actual interventions implemented in the natural setting with high school students. (Academic)

(19) It's another thing **to actually get** involved in a potentially bloody conflict. (Spoken)

Similarly, the component verbs in academic language also convey formality, e.g. *explore*, *navigate*, *engage*, *agree*, while those in spoken English are simple, one-syllable verbs, e.g. *be*, *get*, *say*, *make*, *do*, *have*, *take*. Such a difference between the common verbs specific to each of the particular genres has also been observed by Johansson (2015).

It is evident that the COCA data as of 2019 in the academic texts is in line with data surveyed from 1990-2012 (Johansson, 2015) in that the adverb+verb combinations *to better understand*, *to fully understand*, and *to further explore* are still in the top four, and the results also show that the combinations *to just be*, *to really get*, *to really be*, and *to just say* are in the top four in both Johansson's study (1990-2012) and the latest spoken genre of COCA.

Table 5 Functions of Adverb Splitters in Academic Texts, Spoken English, and TV/Movies

Functions	Academic Text		Functions	Spoken English		Functions	TV/Movies	
	adverb splitter	frequency		adverb splitter	frequency		adverb splitter	frequency
1. act as subjunctions of narrow orientation modifying a gradable verb	<i>better</i> (791) <i>fully</i> (29) <i>further</i> (10) <i>strongly</i> (9) <i>actively</i> (9)	848	1. mark completion	<i>really</i> (437) <i>actually</i> (185)	622	1. act with occurrences that can be imagined as taking place sooner or later along a time scale	<i>finally</i> (152) <i>boldly</i> (78)	230
			2. mark an inclusive or exclusive relationship with	<i>just</i> (258) <i>even</i> (9)	267			

			other possible occurrences					
			3. act as subjuncts of narrow orientation modifying a gradable verb	<i>still</i> (28) <i>better</i> (11)	39			

With regard to the functions of adverb splitters, all the adverbs in the academic texts belong to one function, i.e. to act as subjuncts of narrow orientation modifying a gradable verb, i.e. *better*, *fully*, *further*, *strongly*, and *actively*. This function appeared to be the least common in spoken English as represented by *still* and *better*. It is noteworthy that *better* was the most frequent in academic texts (791 tokens) but occurred with the lowest frequency in spoken English (11 tokens), as previously exemplified. The presence of the only single infinitive phrase *to better understand* is characteristic of formal, academic English commonly used in journal articles, which could explain why this split-infinitive structure was the most common in the academic genre. In contrast, the adverbs with the highest frequency in the spoken genre, e.g. *really*, mark completion (622 tokens), as shown in (20), followed by those marking an inclusive or exclusive relationship with other possible occurrences (267 tokens), e.g. *just*, as in (21). Given that English in TV/movies is spoken, the sole function of adverb splitters, i.e. *finally* (152 tokens) and *boldly* (78 tokens) in this COCA genre interact with occurrences that can be imagined as taking place sooner or later along a time scale.

(20) And it can seem hard **to really be** honest about your emotions, but I think a lot of people feel that way. (Spoken)

(21) I don't think we ought **to just say** the states can do whatever they want. (Spoken)

The use of adverb splitters to indicate high, moderate or minimum degrees, e.g.

better, *fully*, *strongly*, is a dominant characteristic of academic English as such lexical bundles, e.g. *to better understand*, *to fully understand*, and *to successfully navigate*, as in (22), are very common formulaic sequences in general research articles or textbooks. On the other hand, speakers of American English tend to use informal adverbs, *really* and *actually*, in marking completion of an action, as in the bundles *to really get*, *to really be*, *to actually get* and *to actually do*, as exemplified in (23). Others, however, orally use the informal adverbs splitter *just* and *even* to mark an inclusive/exclusive relationship with other possible events, as in *to just be*, *to just say*, and *to even be*, as in (24). Furthermore, they often use two specific expressions *to finally meet* and *to boldly go*, as shown above, in TV and movie subtitles in referring to a time scale.

(22) Lovitts (1996) proposed that students needed two types of cognitive maps **to successfully navigate** doctoral programs, Global CMD and Local CMD. (Academic)

(23) It's another thing **to actually have** the people who are being hurt by these tariffs tell their own stories. (Spoken)

(24) I mean, I was just furious, and I found it easier not **to even be** around a lot of people. (Spoken)

5. Conclusion

This study explored the split infinitives in American English represented in COCA. Of all the genres in COCA, blogs and webpages characterizing informal, written web-based English, had the highest frequency

of split infinitives most likely because they share a similar nature and some of the webpage and blog data overlap (Davies, 2020). Both genres, having been added to the most updated version of COCA, contain the highest number of split infinitives. This demonstrates that the language widely used in webpage/blog writing is of a high level of informality. The main function of the splitters concerns marking an inclusive or exclusive relationship with other possible occurrences. The findings of this study reflect a more realistic use of American-English split infinitives, once assumed to be largely associated with spoken English (Carter & McCarthy, 2006). In addition, it was not unusual to find a number of split infinitives in the two spoken-oriented genres, i.e. spoken and TV/Movies, despite the fact that those in the TV/movies were lower in frequency. In comparison to the academic English in COCA, in which only six different infinitive structures occurred, the spoken genre comprises a wider variety of adverb+verb combinations. While the most important function of adverb splitters in spoken English involves marking completion, those in academic English modify a gradable verb. Overall, the steady rise of split infinitives in American English indicates that this grammatical structure, once considered ill-formed or incorrect, has become more acceptable in not only spoken but also academic English (Johansson, 2015). Put simply, any attempt to prescribe correct usage and proscribe incorrect usage is being challenged, and "...it seems that the split infinitive has been losing its power to shock..." (Leech, Hundt, Mair & Smith, 2009, p. 263). To conclude, grammarians no longer have their absolute right to force English users to conform to long-standing prescriptive traditions when it comes to infinitive splitting (Crystal, 2004).

The study is not without limitations, some of which are worth mentioning here. First, only single-adverb splitters, as opposed to the multiple-adverb splitters (e.g. *to not only* or *to sort of*) and the negative splitters (i.e. *to not*) were investigated in the current study. Future researchers can also analyze these splitters so that they will be able to obtain a clearer picture of the way split infinitives are employed in English (Mikulova, 2011). As for

the second limitation, this study concentrated on one major variety of native-speaker English, i.e. American English. It will be very interesting to look at split infinitives in other Englishes in the Inner Circle, e.g. British English, Australian English, or New Zealand English, and in the Outer Circles, where English is spoken as a second language, e.g. India, Singapore, Malaysia, the Philippines (Calle-Martín & Romero-Barranco, 2014; Gonzalez & Dita, 2016)). A comparative study of this infinitive construction occurring in native-speaker and learner English is also recommended (Phoocharoenkil, 2013). Another limitation of this study is a lack of teachers' and students' attitudes towards infinitive splitting. Further studies can shed light on split infinitives from the viewpoint of English users by taking into consideration teachers' and students' perceptions of using English split infinitives in different contexts (Balla, 2019), e.g. academic and/or weblog writing, EFL textbooks, everyday conversation, etc. The teachers' opinions about whether and how this structure should be introduced in class can also be another interesting research topic (Ebner, 2018; Mitrasca, 2009).

References

- Albakry, M. A. (2005). *Style in American newspaper language: Use and usage*. Unpublished doctoral dissertation, Northern Arizona University.
- Balla, A. (2019). Language Change and Grammar Teaching Books in EFL. *Journal of Contemporary Education Theory & Research*, 3(1), 15-20.
<https://doi.org/10.5281/zenodo.3598688>
- Calle-Martín, J. (2015). The split infinitive in Middle English. *North-Western European Language Evolution*, 68(2), 227-250.
- Calle-Martín, J. & Miranda-García, A. (2009). On the use of split infinitives in English. In A. Renouf & A. Kehoe (Eds.), *Corpus linguistics: Refinements and reassessments* (pp. 347-364). Rodopi.
- Calle-Martín, J. & Romero-Barranco, J. (2014). On the use of the split infinitive in the Asian varieties of English. *Nordic Journal of English Studies*, 13(1), 130-47.

- Carter, R. & McCarthy, M. (2006). *Cambridge grammar of English*. Cambridge University Press.
- Close, R. A. (1987). Notes on the split infinitive. *Journal of English Linguistics*, 20, 217-229.
- Crystal, D. (2003). *The Cambridge encyclopedia of the English language*. Cambridge university Press.
- Crystal, D. (2004). *Rediscover grammar*. Pearson Education
- Davies, M. (2020). *Corpus of contemporary American English*.
- Ebner, C. (2018). Great Britain and the United States: Two nations divided by an attitude? *English Today*, 34(Special Issue 4), 21-28.
- Fitzmaurice, S. (2000). Remarks on the grammaticalization of infinitival to in present-day American English. In O. Fischer and D. Stein (Eds.), *Pathways of change: Grammaticalization processes in older English* (pp. 171-186). John Benjamins.
- Gonzales, W. D. W. & Dita, S.N. (2016). On split infinitives in world Englishes. Paper presented at the 10th international conference of the Asian Association for Lexicography(AsiaLex2016), Century Park Hotel, Manila, 1-3 June, 2016.
- Huddleston, R. D. & Pullum, G. K. (2002). *The Cambridge grammar of the English language*. Cambridge University Press.
- Johansson, S. (2015). *Thou shalt not split...? A corpus-based study on split infinitives in American English*. [Independent thesis, Jönköping University]. Digitala Vetenskapliga Arkivet.
- Leech, G., Hundt, M., Mair, C., and Smith, N. (2009). *Change in contemporary English. A grammatical study*. Cambridge University Press.
- Mikulova, H. (2011). *Split infinitive – Corpus analysis*. [master thesis, Univerzita Palackého Olomouci].
- Mitrasca, M. (2009). The split infinitive in electronic corpora: Should there be a rule? *Concordia Working Papers in Applied Linguistics*, 2, 99-131.
- Peters, P. (2006). English usage: Prescription and description. In B. Aarts & A. McMahon (Eds.), *The handbook of English linguistics* (pp. 759-780). Blackwell.
- Schwarz, S. & Smitterberg, E. (2020). “If anyone would have told me, I would have not believed it.”: Using corpora to question assumptions about spoken vs. written grammar in EFL grammars and other normative works. In E. Jonsson & T. Larsson (Eds.), *Voices past and present – Studies of involved, speech-related and spoken texts* (pp. 283-300). John Benjamins Publishing Company.
- Strumpf, M. & Douglas, A. (2004). *The grammar bible*. Owl Books.
- Torres, J. M. (2019). Positioning Philippine English Grammar and Lexicon in Four Discourse Quadrants. *Asian EFL Journal*.
- Trask, R. L. (2007). *Language and linguistics. The key concepts*. Routledge.