Kennesaw State University

From the SelectedWorks of Chris C. Palmer

2015

The Phrasal Verb in American English: Using Corpora to Track Down Historical Trends in Particle Distribution, Register Variation, and Noun Collocations

David West Brown
Chris C. Palmer

Available at: https://works.bepress.com/chris_c_palmer/2/

David West Brown and Chris C. Palmer

The phrasal verb in American English: Using corpora to track down historical trends in particle distribution, register variation, and noun collocations

1 Introduction

Among their varied findings, corpus-based studies of phrasal verbs in English (such as *run up* in *They always run up our electric bill*) have pointed to two intriguing patterns. First, they have described phrasal verb distributions across registers in Present-Day English, showing their greater propensity for being realized in fiction and spoken text-types (e.g., Biber et al. 1999; Liu 2011). Second, scholars have suggested that the use of phrasal verbs has increased over time (e.g., Spasov 1966; Pelli 1976; Martin 1990; Smitterberg 2008). Our study builds from and draws connections between these two lines of research by combining diachronic and register-based approaches to analyze the development of the phrasal verb in American English. In order to carry out the analysis, this study uses a variety of data sources: *The corpus of historical American English* (COHA) and *The corpus of contemporary American English* (COCA), which are large, monitor corpora, each containing over 400 million words (Davies 2008–; Davies 2010–); qualitative examples of usage from historical archives; and historical records of the production and consumption of material goods.

The investigation comprises two parts. In the first, we present an overview of the development of phrasal verbs in the last two hundred years. Our analysis begins with historical trends in the use of the common adverbial particles and then examines diachronic changes in the frequencies of those particles across registers. In the second part of our investigation, we present case studies of the transitive phrasal verb *pick up* and selected noun collocates. These case studies explore how the use of phrasal verbs is influenced by the histories of the nouns that function as their objects (how the noun *phone* impacts the use of *pick up*, for example). One set of data, therefore, provides a global view of developmental trends focusing on the adverbial particles, the other a close look at some of the grammatical, semantic, and sociolinguistic forces that can influence the histories of specific phrasal verbs.

Overall, our study aims to contribute to the wider body of research on phrasal verbs, as well as more specific conversations about the diachronic development

of phrasal verbs in American English. Furthermore, it places such data on phrasal verbs within broader theoretical and methodological contexts. Thus, this article aims not only to address the possibilities and complications of researching multi-word units, but also to expand the methodological possibilities of analyzing them via corpus-based studies.

Most specifically, our paper argues that, in order to create a full account of the growth or decline of a linguistic form, it is necessary to examine large, continuous, multi-register corpora, rather than to rely solely on an aggregation and comparison of disparate frequency-based studies of that form drawn from corpora of different sizes and registers. By investigating phrasal verbs within large corpora such as COHA and COCA, we are able to push existing quantitative approaches in several underexamined and fruitful directions: (1) the effect of register on the history of phrasal verbs and their particles; (2) the correlations between changes in the use of an individual phrasal verb and changes in nominal objects that frequently collocate with that verb; and (3) the impact of historical change in culture and technology on diachronic changes in phrasal verbs. Moreover, in order to carry out our investigations, we bring data from the material world into conversation with linguistic data, and probe their statistical relationships. These methods lead to several interesting conclusions. One particular register, fiction, contributes to the bulk of phrasal verb growth in nineteenth- and twentieth-century American English. Across multiple registers, verbs accompanied by particles *up* and *out* are the most frequently used phrasal verb forms. And the growth of the most frequent phrasal verb with *up* – i.e., *pick up* – can be explained, in part, by changes in American fashion, transportation, and communication technologies.

2 Diachronic studies of the growth of phrasal verbs in English

Previous quantitative studies of diachronic changes in phrasal verbs, which have focused primarily on frequencies, have shown a general increase in phrasal verb usage from Middle English to Present-Day English. Martin (1990: 101–102) samples letters, from the Paston collection up to private collections in the late twentieth century, illustrating a change from approximately 2000 phrasal verb tokens per million words[1] in the period 1440–1479 to approximately 10,900 tokens per million in 1975–1987. Spasov (1966: 19–21) provides a study of phrasal verb

1 Unless otherwise noted, all token frequencies cited from other studies or from our own data represent normalized counts of tokens of phrasal verbs per one million words of total text.

frequencies relative to verb frequencies in British plays from the late medieval period to 1958. Looking only at dialogue, he argues that the percentage of verbs that are phrasal "shows an almost uninterrupted increase"[2] from the medieval period to the modern – from 3.80% in Middle English, to 5.25% in Shakespeare, to 6.33% in the early eighteenth century, to 12.86% in the post-World War II era (1966: 18). Pelli (1976: 97–127) turns to American English, also examining phrasal verb use in drama. His study similarly shows an overall increase in phrasal verb usage from the late eighteenth century (e.g., approximately 5853 phrasal verb tokens per million in 1765–1775) to the later twentieth century (e.g., approximately 13,780 tokens per million in 1965–1972) (1976: 98–99). Broadening the range of genres investigated but narrowing the historical focus, Claridge (2000) compares phrasal verb frequencies in *The Lampeter corpus*, a multi-register corpus covering 1640–1740, and in LOB, *The Lancaster-Oslo/Bergen corpus*, covering multiple genres from present-day British English. She discovers 3640 phrasal verb tokens per million in *Lampeter*, but more than twice that number – 8536 tokens per million – in the LOB material (2000: 126–127). Smitterberg (2008) examines progressive and phrasal verbs in British English within the multi-register CONCE, *A corpus of nineteenth-century English*. Contrasting the periods 1800–1830 to 1870–1900, he, too, finds an overall increase in phrasal verb use, from approximately 5100 tokens per million at the beginning of the century to approximately 6800 per million at its end (2008: 275).

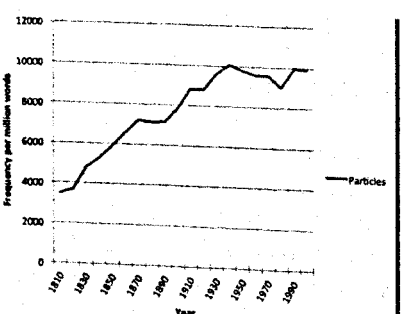
In order to build upon these previous studies, we use COHA, a four-hundred-million-word corpus covering the period 1810–2009, as our primary source of data. The corpus is tagged for part of speech (including for verbs and the particles of phrasal verbs). It also covers every decade within this period, with ample data from a number of distinct registers: fiction, newspapers, magazines, and non-fiction books. The sheer size, range, and continuity of COHA allow for investigations into phrasal verbs that would be impossible to conduct in studies of small corpora. The first of these investigations occurs in Section 3, in which we examine the development of phrasal verbs within individual registers from decade to decade over the last two centuries of American English. In Section 4, the history of some of the most frequent particles is explored. And Section 5 illustrates the ways in which collocation patterns for the particle *up* reveal its semantic history and motivations for use.

To search for verbs, particles, and phrasal verbs in COHA, we followed the methods outlined in Liu's (2011) study of phrasal verbs in COCA, which uses the

2 According to Spasov's charts and graph (1966: 20–21), there is a notable decrease in phrasal verb use to 4.33% of total verbs during the second half of the eighteenth century. By the end of the nineteenth century, he finds a significant increase to 8.88%.

same tagger (CLAWS) as COHA. Verbs were extracted from the corpus by looking for the tag [v*], and particles with the tag [rp*]. Searches for individual phrasal verb constructions considered the tag [v*] [rp*] (e.g., *pick up*), [v*] * [rp*] (e.g., *pick him up*), and [v*] * * [rp*] (e.g., *pick the guy up*). Further separation between the verb and particle was not targeted in searches, since this pattern is infrequent and apt to return false phrasal verbs (Liu 2011: 665).

Figure 1 provides the frequencies of particles used in phrasal verb constructions decade by decade in COHA, while Figure 2 provides the total number of verbs (those appearing in both phrasal and non-phrasal constructions). It is immediately clear from these two charts that the COHA data corroborate the general trends observed in previous diachronic studies: namely, that phrasal verb constructions have largely been increasing in use over time, including for most of the last two centuries in American English. To determine whether or not changes in particle use were primarily dependent on overall changes in verb frequencies, we compared the trends in verbs with trends in the particles (which combined with a subset of those verbs to create phrasal verb constructions). The diachronic distributions of particles and verbs are correlated; as measured by Kendall's-τ (Kendall 1938), the correlation coefficient is 0.786 ($p = 0.0055$).[3] While the Kendall's-τ suggests a significant overall trend, it does not identify or explain any stages or differences in the underlying distributions. In order to further explore the relationship between particles and verbs, we follow the methods set forth in Hilpert and Gries (2009). Specifically, we use variability-based neighbor clustering (VNC), which partitions diachronic data into statistically meaningful sub-periods (Hilpert and Gries 2012). The results of that clustering are plotted in Figures 3 and 4.



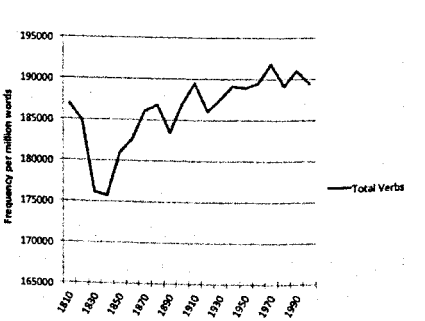
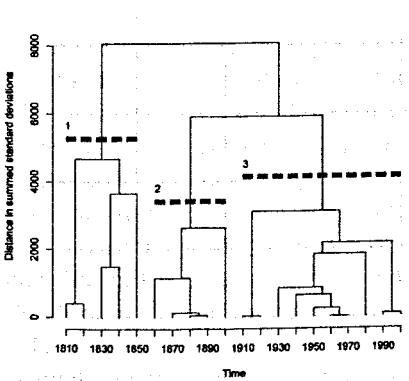


Figure 1 (Left): Token frequencies of particles appearing in phrasal verb constructions, 1810–2009, COHA **Figure 2** (Right): Token frequencies of all verbs, 1810–2009, COHA

3 All statistical analysis for this study was carried out using R (R Core Team 2013).

The dendrogram in Figure 3 shows particles separating into three clusters (as highlighted by the numbered, dashed lines): 1810–1859, 1860–1909, and 1910–2009. Alternatively, the dendrogram in Figure 4 shows verbs separating into five clusters: 1810–1829, 1830–1849, 1850–1869, 1870–1899, and 1900–2009. Thus, although there is a general correlation between the diachronic increase in particles and verbs overall, they do not grow entirely in parallel. The nineteenth century is particularly distinct in this regard. It sees changes in verb frequencies clustered into four statistically meaningful sub-periods, and particles into only two, showing their early histories in American English to be characterized by more volatility for verbs than for particles.

The differences in the growth of particles and the increases in total verbs are further illustrated in Figure 5, which presents the ratio of frequencies of particles to frequencies of total verbs within each decade of COHA. Between 1810 and 1940, even as the use of verbs generally increased, the use of particles in phrasal verb constructions also increased. In other words, the percentage of verbs combining with particles to create phrasal verbs rose steadily during this period, from less than 2% of verbs in 1810 to over 5% of verbs in 1940. It is also worth noting that this ratio has plateaued since 1940, with a small decrease in the ratio of particles to total verbs between the 1940s and the 1980s and a slight rise within the last two decades. These data somewhat corroborate the findings of Pelli (1976: 98–99), who notes a similar trend in particle frequencies within American drama: after steadily increasing in use from the early nineteenth century to the 1925–1935 sub-period, his next sub-period shows a decrease – from roughly 16,000 tokens in 1925–1935 to about 13,780 tokens in 1965–1972. Unfortunately, Pelli's study ends there, so further comparisons



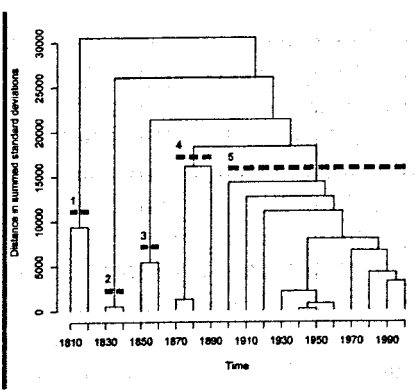


Figure 3 (Left): VNC dendrogram for particles appearing in phrasal verb constructions, 1810–2009, COHA **Figure 4** (Right): VNC dendrogram for all verbs, 1810–2009, COHA

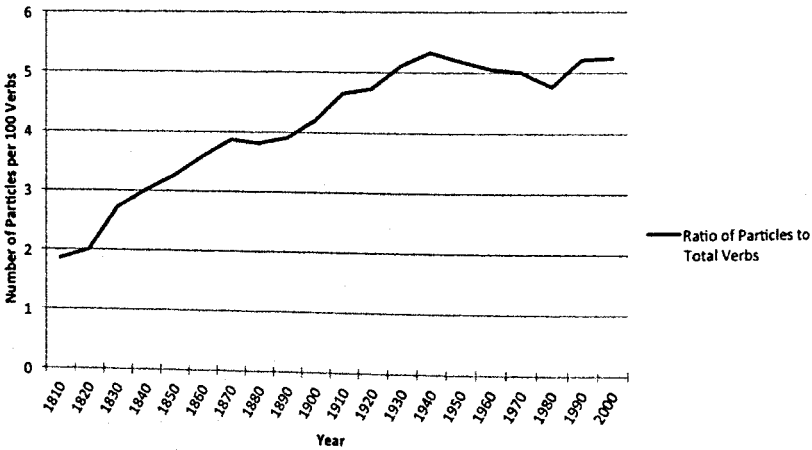


Figure 5: Ratio of token frequencies of particles in phrasal verb constructions to frequencies of total verbs, normalized to occurrences of particles for every 100 verbs, 1810–2009, COHA

in recent decades cannot be made. Even so, the data from COHA provide a more comprehensive picture of the overall diachronic changes in verbs and particles from the last two centuries, nuancing our understanding of the contours of change. While it is clear that phrasal verbs have been on the rise in American English over the last two centuries, that growth seems to have leveled off in the 1940s.

3 Phrasal verbs and register variation in Present-Day English and the history of American English

As the discussion in the previous section makes clear, it is necessary to nuance broad diachronic claims about the growth of phrasal verbs in English. While many studies, including the present one, have generally shown upward trends in the use of phrasal verbs and particles over the centuries, the data from COHA suggest that such growth plateaued in American English beginning in the 1940s. Indeed, Thim (2012: 211–214) cautions that it may not be wise to assume there is consistent growth in phrasal verbs diachronically, especially if frequency studies are used as a sole indicator of growth. Citing a number of frequency studies from different eras (2012: 213), he notes that there is certainly not a steady increase from Middle English to Present-Day English. For example, Castillo (1994: 439–450) finds a frequency of 6500 phrasal verb tokens per million words in Shakespeare; Smitterberg (2008: 275) discovers an average of 5900 tokens in multiple registers of nineteenth-century English; and Biber et al. (1999: 408–409) reports roughly 1400 phrasal verb tokens in Present-Day English within the multi-register *Longman spoken*

and written English corpus. These studies alone would suggest a sharp decline in phrasal verb usage, and run counter to the narrative suggested by Spasov (1966), Pelli (1976), and Martin (1990). Thim's main claim is that frequency largely depends on the type of study conducted on individual registers at particular points in time. He argues that, rather than relying on "general patterns of growth" to explain diachronic change in phrasal verbs, scholars should look deeper into "the development of aspectual and idiomatic constructions over time" and into "changing frequencies in individual text types rather than an overall increase of the construction type" (2012: 214). While Section 5 considers historical changes in idiomatic constructions containing phrasal verbs, the present section digs deeper into the relationship of text-type and diachronic change in phrasal verbs.

Studies of the interplay between phrasal verbs and register in Present-Day English indicate that phrasal verb frequency may be a reliable marker of text-type. Dempsey, McCarthy, and McNamara (2007: 221), for example, argue that phrasal verb frequency correlates strongly with the spokenness and informality dimensions of the spoken/written and informal/formal clines used to classify genres in studies such as Biber (1988). In Biber et al.'s corpus study of spoken and written English (1999: 409), phrasal verbs were found to be most frequent in the registers of conversation and fiction, somewhat frequent in newspapers, and least frequent in academic writing. Liu (2011: 674) presents cross-register data on phrasal verbs from COCA, which we re-present in graphic form (see Figure 6). The

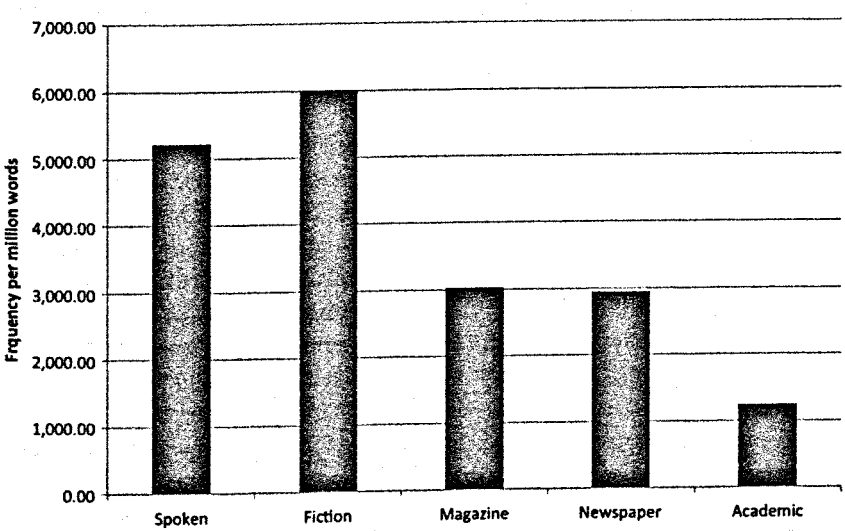


Figure 6: Graphic representation of cross-register variation of phrasal verbs in COCA, normalized to occurrences per million words, from Liu (2011)

results in Liu (2011) corroborate the other studies of Present-Day English phrasal verbs: they tend to occur most frequently in the more oral registers of fiction and speech; they have moderate frequencies in magazines and newspapers; and they have the lowest frequencies in academic prose. But can this sort of distribution be observed in earlier historical periods?

Smitterberg (2008) provides the only broad historical study of register variation and phrasal verbs within late modern English. He finds that in nineteenth-century British English, the registers of drama, fiction, and letters show the greatest increases in phrasal verb frequencies; debates and history show possible moderate increases; and trials and scientific texts show little to no increases (2008: 276–277). While his analysis follows a line of reasoning similar to that of other studies – namely, that the more colloquial registers are most likely to show increases in phrasal verb usage – it is difficult to compare his results, which concern British English only as late as 1900, to those of Liu (2011) on contemporary American English.

COHA, however, provides data on American English up to 2009 for four registers (fiction, magazines, newspapers, and non-fiction), three of which also appear in Liu's (2011) analysis. Hence, by looking at distributions of phrasal verbs among the registers of COHA, we can compare the past and present of American English, determining whether or not phrasal verbs were distinguishing markers of genre well before the present day. Figure 7 provides a historical distribution of phrasal verbs by register. Broadly speaking, the historical data from COHA reflect register patterns observed in the present day: the most speech-like genre within the

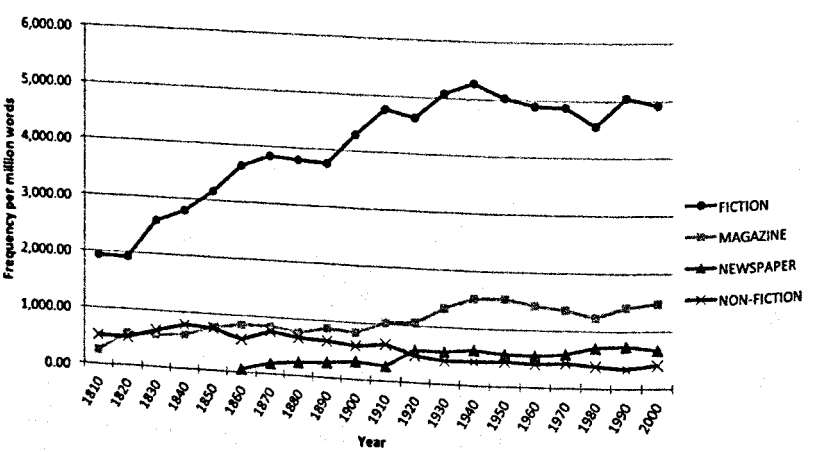


Figure 7: Cross-register variation of phrasal verbs in COHA, normalized to occurrences per million words. Note that newspaper data do not appear until the 1860s in COHA

corpus – fiction – exhibits far higher frequencies of phrasal verbs compared to other registers in every decade of the nineteenth and twentieth centuries. But it is significant to note that these differences have become more marked over time. In the early nineteenth century, fiction's use of phrasal verbs was already higher than, and yet much closer to, the frequencies observed in magazines and non-fiction compared to what can be observed in the late 1900s. Non-fiction, which includes academic writing, shows a steady decline over time. Newspapers and magazines show general increases during these centuries. Even though Figure 6 shows no significant difference between magazines and newspapers in Present-Day English, the data from COHA in Figure 7 suggest that, historically speaking, magazines have had consistently higher frequencies of phrasal verbs than have newspapers.

One explanation for these changes can perhaps be found in Smitterberg (2008: 286), who argues that increases in phrasal verb frequencies within individual text-types are a marker for the "colloquialization" of that text-type. He defines *colloquialization* as "the process whereby a linguistic feature that occurs more frequently in conversational speech than in writing becomes more common than previously in some written genres" (2008: 271). Smitterberg views colloquialization as a prominent development of nineteenth-century British English, in which fiction, drama, and letters adopt increasingly less formal linguistic features – which he tentatively offers as a reflection of the "informalization and democratization" of British society during the nineteenth and twentieth centuries (2008: 282). However, other genres, such as science, have resisted such colloquialization (2008: 278). If phrasal verbs are colloquial markers, it stands to reason that the data in COHA are showing developments parallel to those observed in the registers of Smitterberg's study. American fiction, like British fiction, is perhaps being increasingly colloquialized by writers during this period, as are (to a lesser extent) magazines and newspapers. Non-fiction, which includes academic and scientific writing, shows more resistance to colloquialization.

Another potential factor affecting phrasal verb frequencies may have been editorial practices, especially those enforcing word limits and prescriptive rules. While every writer in every genre represented in COHA may have had word limits to follow before publication, it is arguably the case that writers of shorter genres – such as newspaper and magazine articles – have had much lower word counts (and thus tighter restrictions) to work with than writers of long fiction. Non-fiction works, which include academic monographs and essay collections, might have faced similar editorial demands on word counts. If so, there may have been more pressure on writers in registers such as magazines, newspapers, and non-fiction to choose one-word verbs (e.g., *choose, select*) rather than phrasal verbs

(e.g., *pick out*).[4] In addition to word-limits, Thim (2012: 235–238) speculates that emerging prescriptive norms in the eighteenth century – namely, rules against stranded and sentence-final prepositions (many of which were confused with adverbial particles) and against monosyllabic idioms (which would include many phrasal verbs) – might have impacted the use of phrasal verbs well into the twentieth century. Thim cites several grammars that tried to regulate these features, and it seems likely that editors of magazines, newspapers, and academic books would be most influenced by such prescriptivism. Perhaps fiction writers, even when meeting their own editorial demands, were facing less prescriptive pressure on their choice of verbs than were journalists and academics. However, without direct evidence of editorial practices targeting phrasal verbs due to word limits or prescriptive attitudes, these proposed explanations for historical register variation must remain speculative.

Overall, the investigation of register variation in COHA confirms that phrasal verbs experienced the largest increase in the most speech-like genre, fiction. There were moderate increases in magazines and newspapers, which mix oral and written features, with a small decrease in the least oral register of COHA, non-fiction. More importantly, these data allow us to further nuance claims about the overall trends observed in Figures 1 and 7: not all registers have contributed equally to the rise in use of phrasal verbs in American English. If anything, the vast majority of increase in use as COHA captures it has been due to fiction, which rises sharply until 1940 and plateaus thereafter.

4 Variation of particles in the history of American English

Few studies have compared historical changes in the use of individual particles in phrasal verb constructions within English. Spasov (1966: 24) provides a table with frequencies of each particle within each sub-period of British English, though no trend lines are analyzed, perhaps due to the small size of his corpus and low frequencies of occurrence for many particles. Even so, *up* seems to be the most frequent particle in each century from Middle English to Present-Day English, with

4 Anecdotal evidence suggests that word counts were historically relevant: An 1897 edition of *The Writer*, an American magazine for aspiring writers edited by William H. Hills, states that "articles should be closely condensed (...) [An] ideal length is about 1000 words" (38). In a 1902 edition of *The Writer*, there is an ad for *Comfort* magazine which creates an incentive for pieces that are one hundred words or fewer, so that publishing two one-hundred-word anecdotes is worth $2, while one longer two-hundred word anecdote is only $1.50 (176). Garrett (2002: 85) reports that in the early nineteenth century, Mary Shelley felt the pressure of word-limits more on her shorter writing (e.g., book reviews) than on her longer fiction.

perhaps *out* and *down* the next most frequent. Martin (1990: 109–111) provides a breakdown of particles within her diachronic corpus of letters, noting that *up* and *out* are the most frequent particles in most centuries. She notes some changes as well: in American English, *out* and *up* show similar frequencies in the eighteenth century, though *out* becomes more frequent than *up* in the twentieth. Pelli (1976: 111–114) provides frequencies for several of the most productive particles within his corpus of American drama, calculating the tokens of each particle type as a percentage of total tokens of phrasal verbs within each sub-period. According to his data, *up* is the most frequent particle in all sub-periods from 1765 to 1972. The particle *out* is the second most frequent in all sub-periods, with *back* always the third most frequent. Little change can be observed in Pelli's data from sub-period to sub-period, with the exception of *back*, which steadily increases from the late eighteenth century to the mid-twentieth century. Similar frequencies in Present-Day British English are recorded by Gardner and Davies (2007: 346), who use the *British national corpus* (BNC) as their source of data. They find *up*, *out*, and *back* to be the three most frequent particles, with *down* nearly as frequent as *back*. Furthermore, they find adverbial particle *out* accounts for a full 97.3% of all occurrences of *out*, and adverbial particle *up* 87.4% of all occurrences.

To investigate a broader range of registers and particles, we examined the individual histories of the "six common adverbial particles" listed in Biber et al. (1999: 413) – *up, out, down, on, in, off* – along with *back*. Figure 8 presents

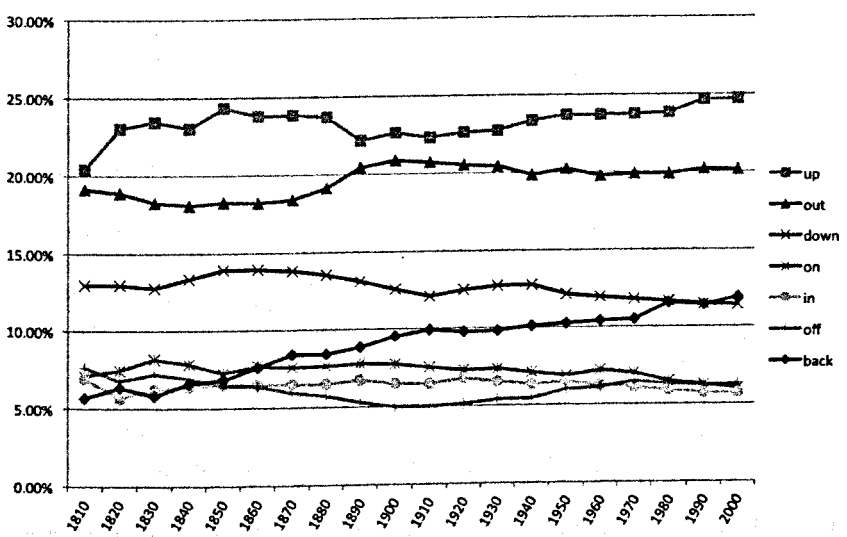


Figure 8: Percentages of selected particles out of the total of all particles in COHA

the trajectories of the various particles as percentages of total particles. In combination with the data presented in Figures 5 and 7, Figure 8 shows that the overall growth in phrasal verbs has been evenly distributed, for the most part, among the most common particles. The particle *up* shows a modest increase in the twentieth century, accounting for a little more than 22% of all particles at the beginning of the century and nearly 25% by the end. Other fluctuations are relatively minor. The exception is *back*, which more than doubles its percentage, going from 5.68% to 11.92%. In fact, *back* surpasses *down* at the end of the twentieth century, as *down* has been declining since about 1940.

Within individual registers, these particles tend to follow the broader trends observed for all phrasal verb types in Figure 7. Even so, by examining particle trends, some differences can be observed within individual genres. For example, consider Figure 9, which presents data on particle trends within the genre of newspapers. Like the general trends, the growth of particles in newspapers is not marked by radical changes in distributions. Some of the particles follow similar trajectories in the newspaper text-type. The particle *down*, for example, decreases as a percentage of the total after 1940 and is overtaken by *back* near the end of the twentieth century. There are, however, also some register-specific patterns. Compared to Figure 8, in which *up* is consistently more frequent than *out* when taking all registers in the aggregate, in newspapers it is only after the 1940s that *up* begins to lead *out* with any consistency.

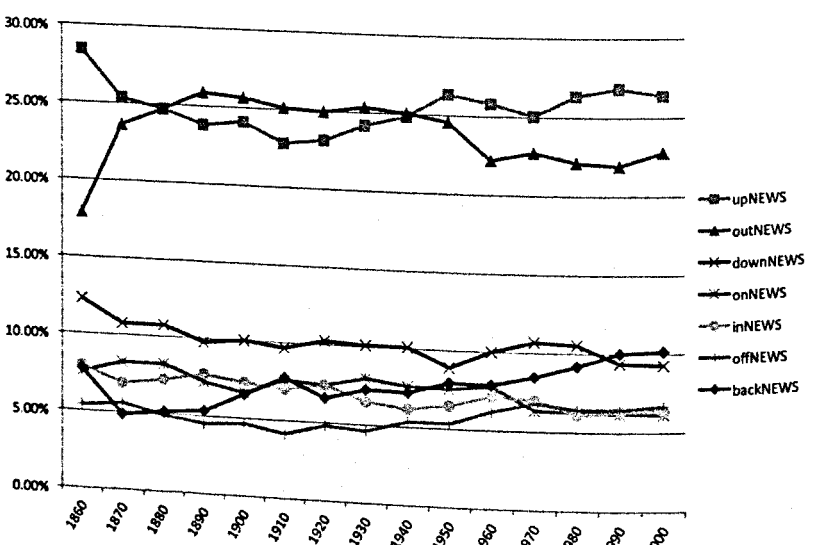


Figure 9: Percentages of selected particles out of the total of particles in the COHA newspaper text-type

An important finding from this investigation is that not all particles contribute equally to the growth of phrasal verbs. It is clear, for example, from Pelli's work on American drama and our own study of multiple registers in American English that *up* has been the largest contributor to the growth of phrasal verbs in American English. In other words, individual particle types seem to have their own individual histories, and those histories may differ somewhat within individual text-types. It is useful for frequency studies to account for these differences among particles to better understand larger patterns of growth among phrasal verbs.

5 A case study of *pick up* and its noun collocates

In an effort to shed further light on trends in usage, we finally want to examine some of the history surrounding the use of a single, common phrasal verb: *pick up*. We choose to focus on this particular combination for a number of reasons. First, as we have noted, *up* is the most common adverbial particle in COHA. Second, *pick* both is the most frequent verbal lemma that collocates immediately to the left of *up* and has the highest association measure as calculated by Mutual Information (see Table 1). Its high MI score suggests not only that the verb *pick* has the greatest affinity for occurring with *up* among possible alternatives, but also that when *pick* appears, it is very likely to co-occur with *up*, as statistically significant association is usually established when MI > 3 (Church and Hanks 1990). By virtue of both its frequency and its degree of association, therefore, *pick up* is a notable and interesting candidate for a case study.

Table 1: The ten most frequent lemmatized verbs that collocate to the left of the adverbial particle *up* in COHA, with their counts, normalized frequencies (per million), percentage of tokens that co-occur with *up*, and association measure as calculated by Mutual Information

Rank	Collocates	Count	Frequency	Percent	MI
1	[PICK] UP	32629	80.32	48.77	8.25
2	[LOOK] UP	31111	76.58	4.97	4.96
3	[COME] UP	25591	63.00	3.03	4.25
4	[MAKE] UP	24042	59.18	2.51	3.97
5	[GET] UP	22750	56.00	3.05	4.26
6	[GIVE] UP	21407	52.70	4.05	4.67
7	[TAKE] UP	20204	49.74	2.76	4.11
8	[SET] UP	16449	40.49	9.50	5.89
9	[GO] UP	14013	34.50	1.31	3.03
10	[STAND] UP	13903	34.22	5.02	4.97

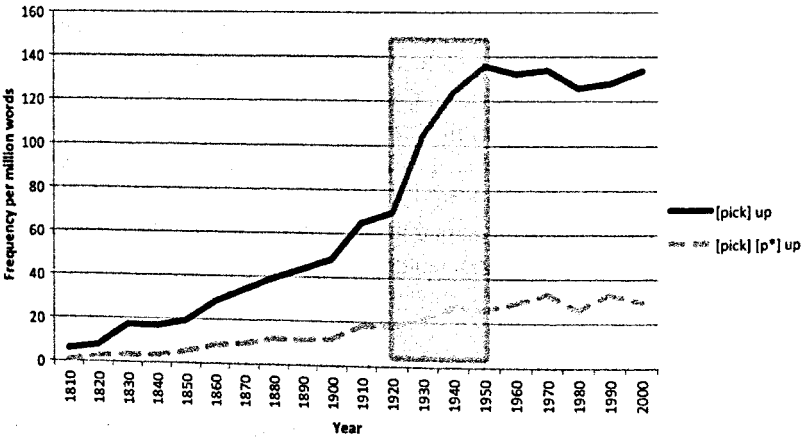


Figure 10: The frequency (per million words) of the lemmatized phrasal verb [*pick*] *up* and [*pick*] *up* with a personal pronoun [p*] between the verb and the particle (e.g., *pick it up*) in COHA

The frequency of *pick up* mirrors trends in the overall growth of phrasal verbs (see Figure 10). Its history is marked by a particularly dramatic increase in the early to middle part of the twentieth century (a period highlighted on the chart). In order to explore some potential reasons motivating this growth, we turn to collocational patterns that emerge during this period. These patterns not only help to explain trends in usage but also shed light on the mutually influential relationship between transitive phrasal verbs and their nominal objects.

Research into V + N combinations has been undertaken from a number of different orientations. One prominent focus in the diachronic study of V + N phrases has been the development of composite predicates (verb + (article) + deverbal noun, e.g., *give a talk*), particularly as such development relates to the processes of lexicalization, grammaticalization, and idiomization (Brinton and Akimoto 1999; Brinton and Traugott 2005; Brinton 2008). V + N collocational patterns have also been the subject of interest in second language learning contexts, where, for example, research has explored the production of such patterns in English learner writing (Nesselhauf 2005; Laufer and Waldman 2007). Taken together, these different areas of research interest and emphasis highlight several important features of V + N combinations: (1) the relationships among verbs and co-occurring nouns are grammatically and semantically structured; (2) co-occurrence can be described probabilistically; (3) V + N relationships change over time; and (4) as Stubbs (1995: 387) puts it, "Culture is encoded not

just in words which are obviously ideologically loaded, but also in combinations of very common words".

While all four of these points are relevant to our analysis here, it is the last two that we want to foreground in framing our discussion. That is, our analysis suggests the complex interactions between cultural changes and increasing phrasal verb frequencies. Specifically, we show how changes in American culture can drive the use of nouns, which, in turn, can drive the use of a phrasal verb through its strong associations with those nouns. In doing so, our analysis further elaborates Rodríguez-Puente's (2012: 83) observation regarding changes in the world and their influence on phrasal verbs. She argues, "Some phrasal verbs have recently been created or have developed new, specific meanings to cope with new technological developments". In this section, we explore how developments in telephone and automobile technologies in the early twentieth century affect the use of *pick up*. Additionally, we investigate a counterexample, a word that has an early association with *pick up*: *hat*. Its decrease in use and association over time seems to correspond to a cultural change, namely, a change in fashion.

Our interest in the connections among *pick up*, its noun collocates, and changes in the material world emerged from the data illustrated in Figure 11 and the posing of a simple question: Does it matter what is being picked up? In other words, are there any diachronic changes in the nouns in V + N phrases that might help us describe changes in the use of the verb itself? Searching for nouns within four words to the right of the verb *pick up* in both COHA and COCA yields the results presented

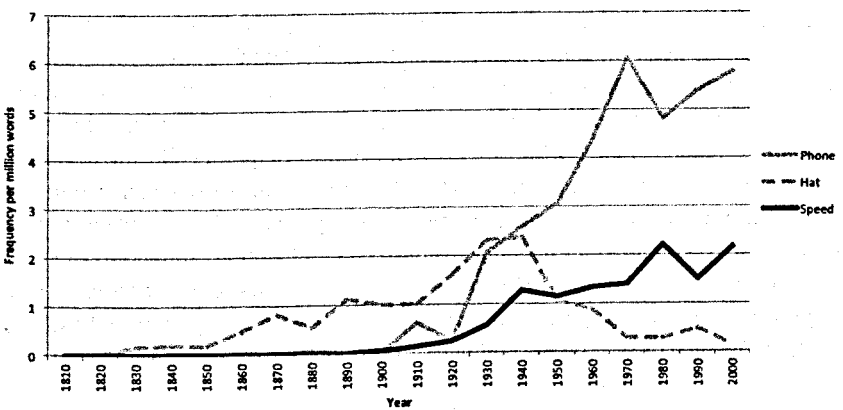


Figure 11: Frequencies (per million) of the lemmatized phrasal verb *pick up* collocating with *phone*, *hat*, and *speed* in COHA

Table 2: The ten most frequent nouns collocating within four tokens to the right of the lemmatized phrasal verb *pick up* in both COHA and COCA

Rank	COHA			COCA		
	Token	Count	Frequency	Token	Count	Frequency
1	PHONE	897	2.21	PHONE	1909	4.11
2	BOOK	363	0.89	SPEED	483	1.04
3	HAT	344	0.85	PIECES	411	0.89
4	SPEED	310	0.76	TAB	383	0.83
5	TELEPHONE	308	0.76	BOOK	359	0.77
6	PAPER	296	0.73	PACE	329	0.71
7	BAG	287	0.71	BAG	249	0.54
8	RECEIVER	260	0.64	PAPER	215	0.46
9	PIECES	239	0.59	SLACK	210	0.45
10	GLASS	223	0.55	RECEIVER	199	0.43

in Table 2. The juxtaposition of the two data sets suggests some interesting points of comparison. The first is that in historical and current American English, *phone* is the most frequent collocate. Additionally, *telephone* and *receiver* are among the ten most frequent collocates in COHA, and *receiver* is among the ten most frequent in COCA. Thus, the lexeme *phone*, along with related derivations and meronyms, appears to have a strong connection to the verb *pick up*. Second, following *phone*, the word *speed* is the next most frequent collocate in current American English and the fourth most frequent historically. In the case of *phone*, the semantic and grammatical relationship between the noun and verb (in a phrase like *pick up the phone*) is fairly straightforward. In the phrase *pick up speed*, however, the meaning of *pick up* has been metaphorized to something like 'gain' or 'increase', and the whole phrase has been idiomaticized. Thus, the motivating factors for the development of *pick up speed* would seem to be different from those for something like *pick up the phone*. Finally, while *hat* is the third most frequent collocate in COHA, it is the sixty-seventh most frequent in COCA. It, then, appears to be an example of change working against the prevailing trend. As opposed to *pick up + phone* or *pick up + speed*, the pattern *pick up + hat* seems to be a collocation falling out of use.

To explain motivations for historical changes in the use of these various combinations with *pick up*, in the following sections we consider external factors, such as changes in technology and material production in American culture.

5.1 *pick up + telephone/phone*

Although *telephone* exists in English as the name for a number of different devices throughout the nineteenth century, the technological developments of Alexander

Bell and Thomas Edison in the 1870s reshape its meaning and spur its proliferation in American English (*OED Online*). This history is confirmed in COHA, where the number of occurrences begins to increase sharply beginning in 1878, one year after Bell and Edison file for patents and the same year the first commercial exchange opens (see, e.g., Fischer 1992). The shortening *phone* appears nearly concurrently; the two forms exist side by side, for example, in this newspaper feature from 1882, with quotation marks calling attention to the shortened form:

(1) *A man was ordering some meat from the market by* ***telephone****, and, after his order was in, he happened to think that he would like some liver, so he put his mouth to the concern and said, "Say, by the way." The connection had been broken, and the attendant at the central office had her ear to the* ***"phone"*** *when the man said, "Have you got any liver?" The girl was nearly frightened to death, but she rallied enough to say, "Why, I suppose so; most girls have a liver. But why do you ask?"* ("An important inquiry" 2)

The lemmatized full form reaches its peak frequency in 1940, when it begins to decline in COHA. The lemmatized shortened form continues to increase in frequency throughout the twentieth century and surpasses *telephone* in COHA in the 1960s (132.08 per million vs. 116.45 per million).[5] Perhaps unsurprisingly, as the technology itself becomes more widespread, so, too, do the words naming that technology. This parallel growth is demonstrated in Figure 12, where the number of telephones in the U.S. (Federal Communications Commission 1946, 1982) is plotted on one y-axis and the frequency of *telephone* and *phone* is plotted on a second. The two trends are highly correlated, with Kendall's-$\tau = 0.956$ ($p < 0.0001$). Thus, there is a strong connection between the-thing-in-the-world and mentions of that thing.

The first co-occurrence of *pick up* + *telephone/phone* does not appear in COHA until 1908 in Upton Sinclair's *The Metropolis*: "And at last, when he did move,

5 The crossover point for *telephone* and *phone* in the Google Books data is 1987 for American English, approximately 20 years later than in COHA. The precise reason for this disparity is difficult to pinpoint. It does, however, appear related to differences in the trends for *telephone*. While *phone* follows similar patterns in the two corpora, the use of *telephone* begins to decline after 1940 in Google Books, as it does in COHA, but then begins to rise again after 1960. It is this rise at the end of the twentieth century that affects the point of intersection. One intriguing observation is that *phone* is most frequent in fiction, while *telephone* is most frequent in newspapers, according to COCA. Thus, the discrepancy may be related to differences in the composition of these corpora, or perhaps due to changes in data made available to Google in more recent decades. In fact, the trends in Google Books using the English Fiction sub-corpus more closely resemble those in COHA.

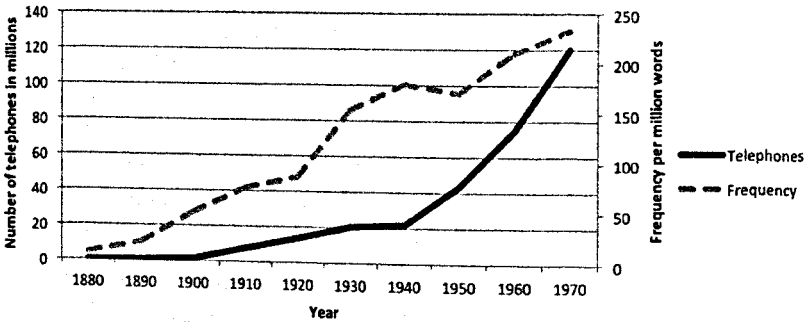


Figure 12: The number of telephones in the U.S. (in millions) plotted against the frequency of the words *phone/telephone* (per million words) in COHA

he **picked up the telephone**, and told his secretary to call up Mr. Hasbrook". Examples of the collocation, however, can be found even earlier. An article in *Popular Science Monthly* (Shaw 1878: 490) describes Thomas Edison in his then-new research facility in Menlo Park, New Jersey: "The Associated Press wires run through his laboratory, and anon **he picks up his telephone** and chats with Philadelphia". This example suggests an early association between these constituent units that is borne out in the data. When Mutual Information is plotted over time, the association between *pick up* and *phone* remains relatively steady throughout the twentieth century, with a slight decline beginning in the 1970s (see Figure 13). While it is not plotted on the graph, the association between *pick up* and *telephone* follows a similar pattern, though the MI scores for *pick up + telephone* are consistently lower than *pick up + phone*. Thus, *pick up* seems to have a preference for the shorter form. The scores for both forms, however, range above three for most of the century. In the case of *phone*, as Figure 13 shows, those scores are significantly higher.

From the combined data, then, a pattern begins to emerge. Unsurprisingly, as telephone technologies spread during the twentieth century, so, too, do the nouns that name the technology. But more interestingly, those nouns do not exist in discourse a-contextually: they have probabilistic associations with other words, one of those being the phrasal verb *pick up*. The increase in the nouns' frequency, therefore, coincides with an increase in the frequency of the phrasal verb.

5.2 *pick up + speed*

While the noun in the phrase *pick up + phone* impacts frequencies of the co-occurring verb, the relationship between noun and verb is somewhat different in

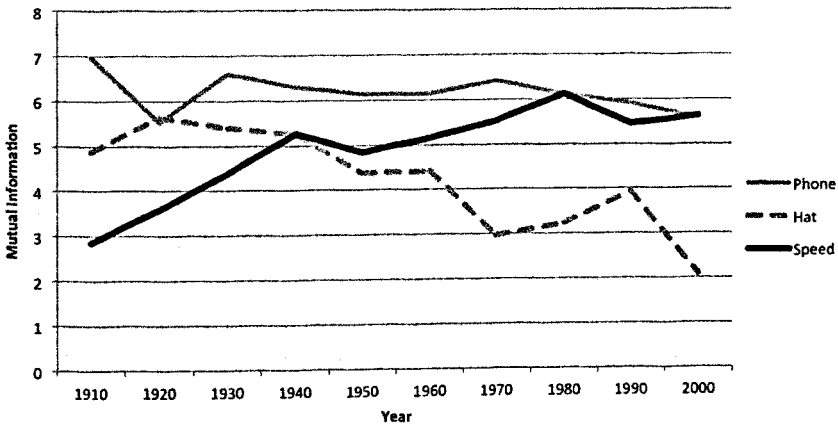


Figure 13: Mutual Information measuring the association between the phrasal verb *pick up* and the nouns *phone*, *hat*, and *speed* plotted from 1910 to 2000

the case of *pick up* + *speed*. Although the use of the phrase *pick up speed* increases (with a slight dip after 1980) after it first appears in COHA in 1911, the use of *speed* as a noun remains relatively consistent over the course of the twentieth century. The frequencies of the idiomaticized *pick up speed* do not have any obvious relationship to the frequencies of its noun constituent.

In order to contextualize changes in *pick up* + *speed*, we look instead to the nouns that precede the phrase. In both COHA and COCA, the three most common collocates (searching 6 to the left) are *train*, *car*, and *plane*. In COHA, all three have scores where MI > 3 with a minimum of 5 occurrences (*train* MI = 7.14, *plane* MI = 5.88, *car* MI = 5.43). In COCA, only *train* and *car* have scores where MI > 3 (*train* MI = 6.83, *car* MI = 4.14). The phrase *pick up speed* can appear in a variety of contexts, as in the following example from 1916, where it is paired with another idiom ("the ball is rolling"). Its understood subject (something like "plans for American agriculture's economic stability") further metaphorizes the semantic content of the V + N phrase:

(2) *It has been an epoch-making work, for these fundamental steps which have been taken are the enabling acts that will eventually put American agriculture for the first time on a sound economic footing. Not all of the plans that have been devised along this line have been put into actual operation yet, but the ball is rolling and* ***picking up speed*** *from day to day.* (COHA: MAG)

Although such uses are not unusual, the collocational data show that *pick up* + *speed* has a strong association with transportation technologies. That association is attested in examples that predate any data from COHA, like this one from the

Proceedings of the Engineers' Club (1892: 366), which the minutes attribute to comments made by William S. Aldrich, a professor:

(3) *In Chicago the overhead steam roads can hardly keep up in speed with the cable surface roads. The great advantage of the cable is that the car, after stopping, very soon* ***picks up speed*** *– in Philadelphia, indeed, almost too quickly. A steam train cannot do this. The cable gives the most rapid surface transit possible. After it comes steam, and then electricity. The storage battery is best on crowded streets, and the trolley on long stretches of open streets and park. It can* ***pick up speed*** *more quickly than the storage battery car.*

Notably, *pick up + speed* emerges in the U.S. around the same time as *pick up + telephone/phone* – in the latter part of the nineteenth century. It has a strong collocational association with rail transportation, which has a different trajectory in American discourse (the use of the noun *train*, for example, gains in frequency through the nineteenth century, peaks in the 1920s, then declines). This collocational pattern likely motivates some of the early growth of the idiom. However, the idiom's other significant collocate, *car*, also influences its history. In Figure 14, the number of automobiles registered in the U.S. (Department of Transportation 2011) is plotted against the frequency of the words *car/automobile*. Like the word-object data for *telephone/phone*, the Kendall's-τ for the data in Figure 14 suggests a significant correlation: $\tau = 0.642$ ($p < 0.0001$). The somewhat weaker correlation and differences in the tailing left ends of the charts can probably be attributed to two factors. First, automobile registration was not required early on in most states, so the number of registrations between 1900 and 1910 does not accurately reflect the number of cars. Second, as the quotation above illustrates, *car* does not necessarily refer to an automobile, but may refer to a trolley, train, or some other passenger compartment. Thus, the frequency is capturing more than references

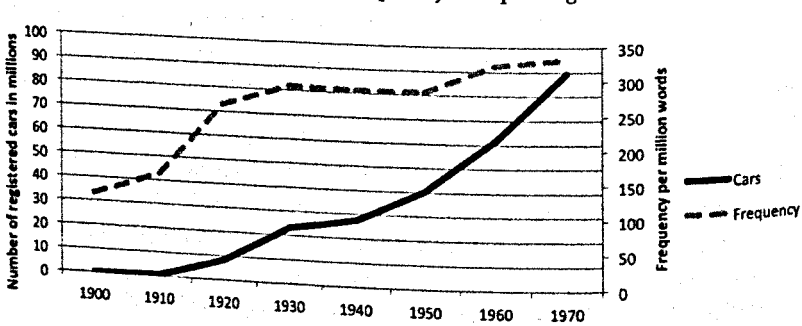


Figure 14: The number of automobiles registered in the U.S. (in millions) plotted against the frequency of *car/automobile* (per million words) in COHA

to automobiles. This range of meaning causes greater distortion early (when there are competing forms) than late (when *car* as automobile is the dominant form). For example, from 1910 to 1919 there are more bigrams of *street car* (n = 81) than *motor car* (n = 53). From 2000 to 2009, common bigrams include *police car* (n = 88) and *rental car* (n = 83), and there are more instances of *rent-a-cop car* (n = 28) than *subway car* (n = 17) or *dining car* (n = 15).

The data, then, seem to indicate a history of *pick up + speed* that is somewhat more complex than that of *pick up + telephone/phone*. It is a collocation that develops as an idiom frequently used in association with rail technology, but is also associated with emergent automobile technology. As the latter proliferates, so too does the V + N phrase. But despite the differences in the grammatical and semantic relationships between the phrasal verb and various nouns signifying communication and transportation technologies, in each case the diachronic frequencies of *pick up* seem to be influenced by changes in the material world. It is worth noting, too, that in the U.S. these changes in telecommunications and transportation are happening concurrently. Fischer (1992), for example, remarks on their mirrored growth in the first half of the twentieth century and dual impact on American culture. These profound cultural changes are partly what lie behind the linguistic changes highlighted in Figure 10.

5.3 *pick up + hat*

Finally, we want to briefly discuss an example that runs counter to the prevailing increase in the use of *pick up*: the collocation *pick up + hat*. Unlike *pick up + telephone/phone* and *pick up + speed*, we do not tie *pick up + hat* to changes in technology. However, the rise and fall of hats in American fashion has been interpreted as an indicator of cultural change (e.g., Lieberson 2000; Schudson 1984; Tenner 1989). Tenner (1989: 22), for example, describes images of crowds prior to World War II as "hatscape[s]", and he connects the decline of hats to a decreasing preference for overt marking of social hierarchy. In his summary of some of the arguments advanced about the decline of hats in American fashion, Lieberson (2000: 82) lists claims such as the emergence of hair as a marker of self-expression; President Kennedy's personal (often hatless) style; and the rise of cars, which are often difficult to enter and exit while wearing a hat. For his part, Lieberson (2000: 83) argues that the decline in the popularity of hats began earlier than most explanations assume. Their decreasing popularity was one feature of an overall decline in formality, which includes other forms of self-presentation besides clothing.

In support of his claims, Lieberson presents data for hats and gloves from the Sears, Roebuck catalogue (see Figure 15). These figures capture the offerings

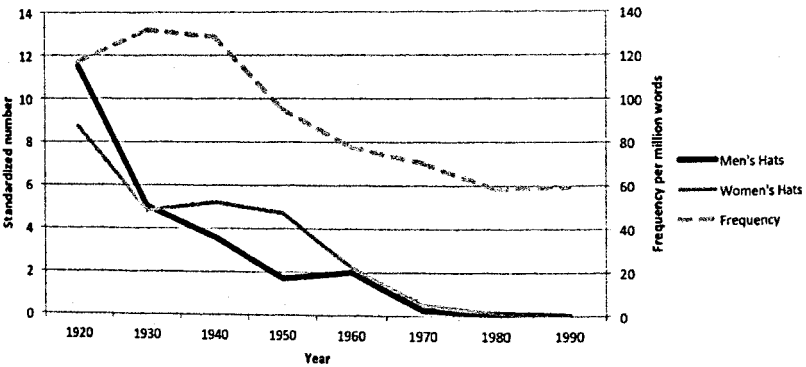


Figure 15: The popularity of men's hats and women's hats in the Sears, Roebuck catalog, 1920–1990, from Lieberson (2000: 79). The standardized number on the y-axis represents an effort to control for the fluctuating size of the catalogue and is calculated by dividing the number of hats (excluding casual wear, winter hats, ski hats, and baseball caps) by the pages devoted to men's underwear. Lieberson's data is plotted against the frequency of *hat* in COHA.

of only one company (and are thus tied to its varying fortunes) and may not always reflect patterns of consumption. Thus, while there is a positive correlation between the word frequency and the standardized number (Kendall's-τ = 0.713, $p = 0.0141$), there are significant limitations to that correlation. Even so, the catalog data intersect with the linguistic data in interesting ways and offer a kind of counterpoint to data presented in Figures 12 and 14. The frequency of the word *hat* peaks in 1910–1919 (133.56 per million) and remains relatively steady through 1940–1949 (128.39 per million), when it begins a steady decline. The collocational pattern of *pick up* + *hat* peaks in 1920–1929, a decade later than *hat* by itself, but otherwise mirrors the trend in use for the single word (see Figure 11). Thus, both *pick up* + *hat* and *hat* are gaining or maintaining frequency when Lieberson's graph shows a decline. One possibility is suggested by the data presented in Olds' (2011) study of the hat industry in Taiwan. Although it is data from outside the U.S., it shows the production of hats that were primarily intended for the U.S. market. His data reveal output rising until 1934, even as the value of that output decreases from the effects of the depression. Lieberson's chart may be reflecting similarly complicated interactions among the demand for a thing, its availability, and its value during a period of economic decline.

In the case of *hat*, the relationship between the circulation in the material world and circulation in language is certainly difficult to quantify in the early part of the twentieth century. However, the downward trends in both are clear after the 1940s. Additionally, *hat* decreases its association with *pick up* over the course of the twentieth century, dropping below the level of statistical significance in 2000

(see Figure 13). This is in contrast to *phone*, which largely maintains its association with *pick up*, and *speed*, which increases its association with *pick up* over time. In this way, *pick up* + *hat* serves as an important caveat. It demonstrates that while some collocational patterns may reinforce prevailing trends in use of a particular constituent, there may be concurrent patterns working as counterweights. *Pick up* may be riding the coattails of nouns related to telecommunication and transportation technologies, but those are not the sole forces driving its diachronic frequencies. This latter point is key. While we demonstrate here the importance of collocational associations and the sometimes surprising connections between word frequencies and changes in the material world, these represent only some of the circumstances contextualizing the history of the phrasal verb *pick up*.

6 Conclusion

On the basis of our study of phrasal verbs in American English, we offer the following considerations and conclusions:

- Frequency studies tell us some aspects of linguistic history. But that history can be a spotty one. In the case of phrasal verbs, studies of individual registers have painted a picture of mostly unimpeded growth from Middle English to Present-Day English. But such studies have made it difficult to make comparisons between eras and genres to evaluate the contours of such growth in the overall language, especially if only small corpora in randomly selected sub-periods are considered. Such findings can be better understood by comparing them with data drawn from larger, continuous, multi-register corpora such as COHA. Specifically, while we have found that phrasal verbs in the aggregate have continued to increase in use in nineteenth- and twentieth-century American English, that growth seems to have plateaued around the 1940s.
- It is important to nuance our understanding of "growth" in phrasal verb use by exploring narrower quantitative developments within broader quantitative accounts – whether we mean "broad" in its synchronic sense ("throughout the language in any one period") or its diachronic one ("over a long period of time"). One way to achieve a more nuanced analysis is to consider changes within registers and changes among different particles. Perhaps confirming the increasing colloquialization of some genres, our investigation of register in American English finds that phrasal verbs have grown in use in fiction, newspapers, and magazines – with fiction being the greatest contributor to overall growth by far – while they have decreased in use in nonfiction. And among particles in American English, the particles *up* and *out*

continue to be some of the most common and increasingly frequent forms in phrasal verb constructions.

- Quantitative data become a richer form of evidence when put into the context of historical changes in culture. To achieve a richer analysis of diachronic change in phrasal verbs, we first used broad quantitative studies to zero in on individual data points worthy of further investigation – in this case, the most frequent particle historically (*up*) and its most frequent collocating verb form (*pick*). We then inspected the most frequent noun collocations of the verb and particle combined. To analyze the history of *pick up* and its nominal collocates, we juxtaposed the frequency of linguistic forms with data on historical changes within the material production of the real-life referents of those forms. In the case of this particular phrasal verb, it became clear that changes in both fashion (with reference to collocate *hat*) and technology (in communication, with reference to the collocate *phone*, and in transportation, with reference to *speed*) impacted changes in the frequency and semantics of *pick up*. While the particular impact of these historical developments cannot automatically be assumed to be relevant and explanatory for all phrasal verbs, our study demonstrates the value of considering such external factors in any detailed study of motivations for historical change for particular constructions.

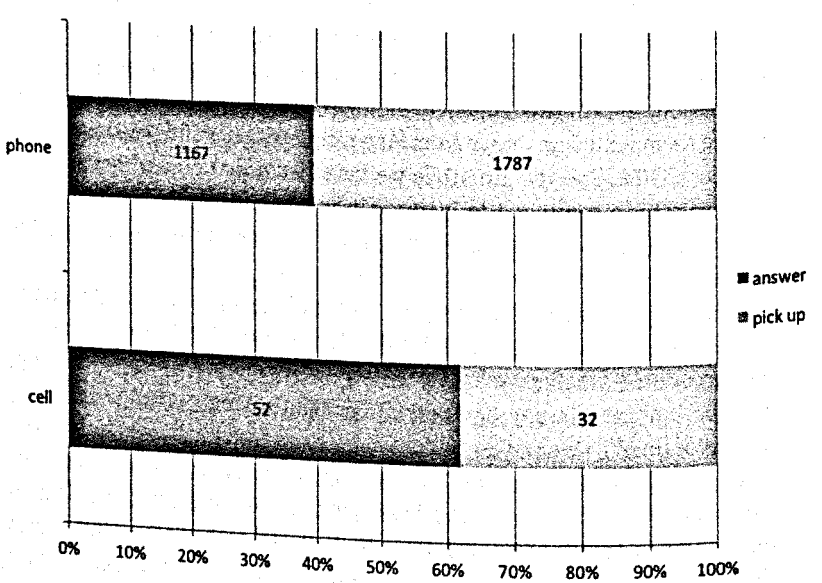


Figure 16: Percentages in COCA of *phone* and *cell* appearing in phrases with lemmatized *answer* and *pick up* of the kind V + (*) + N (e.g., *answers her phone* vs. *picks up the phone*)

In the specific case of *pick up*, it will be interesting to see how, or even if, more recent changes in telecommunications technologies will influence collocational patterns. While we have more phones than ever, we have new words to refer to those phones – notably, in the U.S., *cell phone*, or *cell*. It appears that while *phone* has preference for *pick up* over *answer* (e.g., *picked up the phone*), *cell* or *cell phone* has a preference for *answer* (e.g., *answered the cell*) (see Figure 16). It is possible that such changes may depress future frequencies of *pick up*. However, we will still be *picking up speed*, not to mention *picking up the tab* and *picking up the pieces*. It is just as likely, too, that *pick up* will develop new collocations and new meanings as phrasal verbs are an important, if sometimes obscure, trace of the interaction between the material world and language change.

References

Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman grammar of spoken and written English*. London: Longman.

Brinton, Laurel J. 2008. "Where grammar and lexis meet": Composite predicates in English. In Elena Seoane and María José López-Couso (eds.), in collaboration with Teresa Fanego, *Theoretical and empirical issues in grammaticalization*, 33–53. Amsterdam & Philadelphia: John Benjamins.

Brinton, Laurel J. & Minoji Akimoto (eds.). 1999. *Collocational and idiomatic aspects of composite predicates in the history of English*. Amsterdam & Philadelphia: John Benjamins.

Brinton, Laurel J. & Traugott, Elizabeth Closs. 2005. *Lexicalization and language change*. Cambridge: Cambridge University Press.

Castillo, Concha. 1994. Verb-particle combinations in Shakespearean English. *Neuphilologische Mitteilungen* 95. 439–451.

Church, Kenneth Ward & Patrick Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational Linguistics* 16(1). 22–29.

Claridge, Claudia. 2000. *Multi-word verbs in Early Modern English*. Amsterdam & Atlanta: Rodopi.

Davies, Mark. 2008–. *The corpus of contemporary American English: 450 million words, 1990–present*. http://corpus.byu.edu/coca/ (accessed 1 September 2013).

Davies, Mark. 2010–. *The corpus of historical American English: 400 million words, 1810–2009*. http://corpus.byu.edu/coha/ (accessed 1 September 2013).

Dempsey, Kyle B., Philip M. McCarthy & Danielle S. McNamara. 2007. Using phrasal verbs as an index to distinguish text genres. In David Wilson & Geoff Sutcliffe (eds.), *Proceedings of the twentieth international Florida artificial intelligence research society (FLAIRS) conference*, 217–222. Menlo Park: The AAAI Press.

Department of Transportation, United States. 2011. *Highway statistics summary to 1995*. http://www.fhwa.dot.gov/ohim/summary95/section2.html (accessed 2 October 2013).

The Engineers' Club of Philadelphia. 1892. *Proceedings of the Engineers' Club of Philadelphia,* Vol. 9. Philadelphia: Published by the Club.

Federal Communications Commission, United States. 1946. *Statistics of the communications industry in the United States for the year ended December 31, 1944.* Washington, D.C.: United States Government Printing Office.

Federal Communications Commission, United States. 1982. *Statistics of communications common carriers.* Washington, D.C.: United States Government Printing Office.

Fischer, Claude S. 1992. *America calling: A social history of the telephone to 1940.* Berkeley, CA: University of California Press.

Gardner, Dee & Mark Davies. 2007. Pointing out frequent phrasal verbs: A corpus-based analysis. *TESOL Quarterly* 41(2). 339–359.

Garrett, Martin. 2002. *Mary Shelley.* New York: Oxford University Press.

Hills, William H. 1897. *The Writer: A Monthly Magazine for Literary Workers. Volume X.* Boston: The Writer Publishing Company.

Hills, William H. 1902. *The Writer: A Monthly Magazine for Literary Workers. Volume XV.* Boston: The Writer Publishing Company.

Hilpert, Martin & Stefan Th. Gries. 2009. Assessing frequency changes in multistage diachronic corpora: Applications for historical corpus linguistics and the study of language acquisition. *Literary and Linguistic Computing* 24(4). 385–401.

Hilpert, Martin & Stefan Th. Gries. 2012. Variability-based neighbor clustering: A bottom-up approach to periodization in historical linguistics. In Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford handbook of the history of English*, 100–124. Oxford: Oxford University Press.

An Important Inquiry. 1882, May 28. *Los Angeles Herald.* 2.

Kendall, M. G. 1938. A new measure of rank correlation. *Biometrika* 30. 81–93.

Laufer, Batia & Tina Waldman. 2011. Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning* 61(2). 647–672.

Lieberson Stanley. 2000. *A matter of taste: How names, fashions, and culture change.* New Haven, CT: Yale University Press.

Liu, Dilin. 2011. The most frequently used English phrasal verbs in American and British English: A multi-corpus examination. *TESOL Quarterly* 45(4). 661–688.

Martin, Pamela. 1990. The phrasal verb: Diachronic development in British and American English. New York: Columbia University Ph.D. thesis.

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, William Brockman, The Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak & Erez Lieberman Aiden. 2011. Quantitative analysis of culture using millions of digitized books. *Science* 331(6014). 176–182.

Nesselhauf, Nadja. 2005. *Collocations in a learner corpus.* Amsterdam & Philadelphia: John Benjamins.

Olds, Kelly B. 2011. The Taiwan hat industry: Pre-war roots of the post-war miracle. *Business History* 53(7). 1110–1129.

Pelli, Mario G. 1976. *Verb-particle constructions in American English: A study based on American plays from the end of the 18th century to the present.* Bern: Francke.

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/ (accessed 15 February 2014).

Rodríguez-Puente, Paula. 2012. The development of non-compositional meanings in phrasal verbs: A corpus-based study. *English Studies* 93(1). 71–90.
Schudson, Michael. 1984. *Advertising, the uneasy persuasion: Its dubious impact on American society.* London: Routledge.
Shaw, G. M. 1878. Sketch of Thomas Alva Edison. *Popular Science Monthly* 13(23). 487–491.
Smitterberg, Erik. 2008. The progressive and phrasal verbs: Evidence of colloquialization in nineteenth-century English? In Terttu Nevalainen, Irma Taavitsainen, Päivi Pahta & Minna Korhonen (eds.), *The dynamics of linguistic variation: Corpus evidence on English past and present*, 269–289. Amsterdam & Philadelphia: John Benjamins.
Spasov, Dimiter. 1966. *English phrasal verbs.* Sofia: Naouka i Izkoustvo.
Stubbs, Michael. 1995. Collocations and cultural connotations of common words. *Linguistics in Education* 7(4). 379–390.
Tenner, Edward. 1989. Talking through our hats. *Harvard Magazine* 91. 21–36.
"telephone, n." OED Online. September 2013. Oxford University Press. http://www.oed.com/view/Entry/198718?rskey=GovGzr&result=1 (accessed 2 October 2013).
Thim, Stefan. 2012. *Phrasal verbs: The English verb-particle construction and its history.* Boston & New York: De Gruyter Mouton.