Proceedings of ICADSS 2026 Symposium: Interpreting Culture: Context in Art, Literature, and Mass Media
DOI: 10.54254/2753-7064/2026.HT34545

A Corpus-Based Study of the Decline of Core Modal Verbs in English: Evidence from COHA and COCA

Yihan Sun

School of Political Science and Public Administration, Shandong University, Qingdao, China
202200031061@mail.sdu.edu.en

Abstract. As English continues to change in both spoken and written communication, the use of modal expressions has also shifted. Core modal auxiliaries such as must and shall appear to be less dominant, while semi-modal constructions such as have to have become increasingly common. This study investigates the diachronic change and functional relationship among must, shall, and have to in American English. It aims to examine whether must and shall decline over time, whether have to rises, and whether have to can partially replace must in expressing obligation or necessity. Using COHA and COCA as the main data sources, this study combines frequency analysis, register comparison, Pearson correlation, and semantic coding of concordance lines. The results show that must and shall decline over time, while have to increases significantly, especially in spoken English. The study concludes that have to represents a partial functional replacement of deontic must and reflects broader processes of colloquialization and semi-modalization.

Keywords: modality, semi-modalization, colloquialization, corpus linguistics

1. Introduction

The English modal system has undergone noticeable changes over time. In contemporary English, core modal auxiliaries such as *must* and *shall* have declined, while semi-modal constructions such as *have to* have become increasingly important.

This developmental process has been proven in traditional research, relating to colloquialization and grammatical change [1, 2]. More recent research has also stressed that register and functional meaning play an important role in explaining modal variation [3]. Furthermore, studies on American English argued that the distribution of modals and semi-modals are not even across time and contexts [4].

Under this background, this study aims to answer two research questions, combining diachronic and synchronic perspectives together. First, how has the usage frequency of must, shall, and have to changed over time? Second, to what extent can have to be interpreted as partially replacing deontic must in expressing obligation or necessity? To answer these questions, this study utilizes The Corpus of Historical American English (COHA) and the Corpus of Contemporary American English (COCA), taking advantage of diverse methods, such as analyzing frequency, making comparisons across registers, calculating Pearson correlation, and sorting concordance lines based on their semantic meanings.

DOI: 10.54254/2753-7064/2026.HT34545

2. Methodology

2.1. Corpora

This study adopts a corpus-based approach to examine diachronic change in English modal systems. Corpus linguistics allows researchers to investigate language use through large-scale, computer-assisted analysis of authentic texts, especially in terms of frequency, meaning, and distribution [5].

The study used two corpora. COHA is used for examining diachronic frequency, and COCA is used for register comparison. COHA covers a long historical period, providing a panorama of the changes in modal expressions [6]. COCA includes multiple registers, including spoken English, fiction, magazines, newspapers, and academic writing [7].

Together, the two corpora allow the researcher to conduct a diachronic study of the important changes in the modal system, supplemented by register-based variation from a synchronic perspective.

2.2. Methods

Must, *shall*, and *have to* are the target items of this study. It is widely acknowledged that core modals (*must, shall*) are used to express obligation, necessity, or show formality in tradition. This study argues that semi-modal constructions (*have to*) can overlap with some core modals (*must, shall*), which is especially true in expressing obligation and necessity. Observing the changing trajectory and functional relationship of the three items helps figure out the answers to the research problems.

In terms of data collection, this study obtains the frequency data of *must*, *shall*, and *have to*from the two corpora (normalized per million words). This normalization is the prerequisite of later comparison across decades and registers, avoiding the influence of corpus size.

The results are then analyzed from both quantitative and qualitative perspectives. First, the frequency data depict the developmental trajectory of *must*, *shall*, and *have to*. Second, Pearson correlation provides supplementary statistical evidence to examine to what extent that the decline of *must* and *shall* and the rise of *have to* are related by observing whether the move is occurred together. Thirdly, to observe the overlap in a more precise way, this study adopts systematic sampling and semantic coding to classify *must* in selected concordance lines based on its meaning. Among multiple meanings, the expression of obligation and necessity is particularly important for making a sound claim of partial replacement.

3. Results

3.1. Frequency trends

Figure 1 shows the diachronic frequency trends of *must, shall*, and *have to* in COHA from 1820 to 2010. A clear divergence between the two core modal verbs and the semi-modal *have to* can be observed from the line chart.

DOI: 10.54254/2753-7064/2026.HT34545

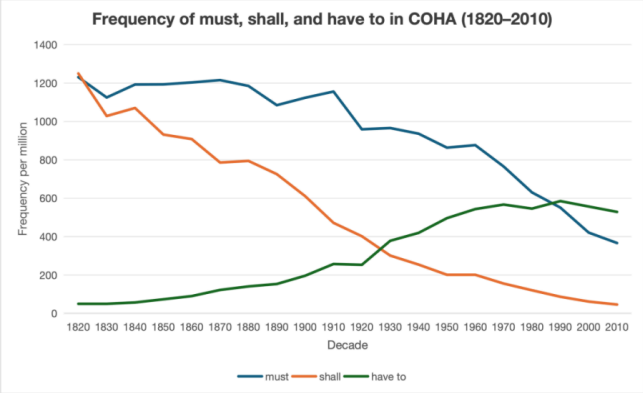

Figure 1. Frequency of must, shall, and have to in COHA (1820-2010)

First, *shall* shows the most dramatic decline. Although *shall* occurred frequently in the early nineteenth century, it started to decline constantly at a fast rate and was much less common by the early twenty-first century. This trajectory shows that *shall* is becoming less and less remarkable in general language use, especially in recent decades.

Second, although the frequency of *must* remained at a high level in the nineteenth century, its clear decline begins from the 1910s. This trend lasts until the end of the period, indicating a long-term downward trend of *must*.

Third, the trajectory of *have to* shows an opposite pattern, whose frequency rises steadily overall. To be more precise, its frequency first surpasses shall around 1930, and later surpasses must around 1990. The two crossover points mark a significant change in the modal system.

Overall, the trends shown in Figure 1 support the view that English modal usage has shifted from some core modal verbs toward semi-modal constructions.

3.2. Replacement patterns

3.2.1. Correlation results

Table 1. Pearson correlations among must, shall, and have to

pair	Correlation
must vs *have to*	r = -0.882
shall vs *have to*	r = -0.962
must vs *shall*	r = 0.854

Pearson's r ranges from -1 to +1. A value close to +1 means similar movement, while a value close to -1 means opposite movement. Table 1 presents the Pearson correlation results among *must, shall*, and *have to* (calculated to three decimal places). Together with the evidence shown in Figure 1, these evidence further support the broader change in modal system.

As seen in Table 1, the correlation between *must* and *have to* is strongly negative, with r = -0.882. The negative relationship is even more marked between *shall* and *have to*, with r = -0.962. These two figures suggest that the increase of the frequency of *have to* may be accompanied by the decrease of the frequency of *must.*

The strong positive correlation between *must* and *shall* (r = 0.854) aligns with the finding in Figure 1 that the two core modals both followed a declining trajectory, moving in the similar

DOI: 10.54254/2753-7064/2026.HT34545

direction.

It is necessary to notice that the correlation can not be viewed as proof of direct causal replacement. It can only support a broader modal-system shift. Namely, the decline of core modals (*must, shall*)and the rise of semi-modal constructions (*have to*) are strongly associated.

3.2.2. Semantic coding

Table 2. Semantic coding results for must in COCA spoken data

Semantic type	Count	Percentage
Deontic	26	52.00%
Epistemic	18	36.00%
Other/ formulaic	6	12.00%
Total	50	100.00%

To avoid overgeneralizing the replacement claim, this study conducted semantic coding of *must*.

The item *must* has multiple meanings. It can express deontic meaning, such as obligation or necessity. It can also express epistemic meaning, such as inference or logical certainty. In addition, some uses are formulaic or non-central, such as I *must* say or nominal uses like a political *must*.

In this study, a total of 200 concordance lines of *must* were retrieved from the spoken section of COCA. Every fourth line was selected, producing a sample of 50 examples. Each example was then manually coded as deontic, epistemic, or other/formulaic.

Table 2 shows the results of this semantic coding. Among the 50 examples, 52% are deontic, 36% are epistemic, and 12% are formulaic or non-central uses. This result is crucial because it indicates that only deontic *must* directly overlaps with *have to* in expressing obligation or necessity, supporting a partial functional replacement claim.

Together, the correlation results and the semantic coding results suggest that the rise of *have to* is not only about frequency increase. It also has functional relevance as *have to* overlaps with *must* in one of its central semantic domains. However, since *must* also has epistemic and formulaic uses, the replacement relationship remains partial.

4. Discussion

4.1. Colloquialization: register-based and pragmatic shift

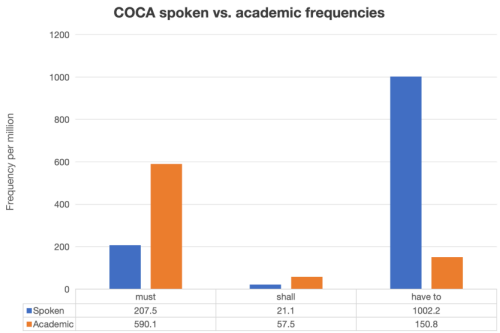

Figure 2. Spoken and academic frequencies of must, shall, and have to in COCA

DOI: 10.54254/2753-7064/2026.HT34545

The spoken-academic comparison strongly supports that *must*, *shall*, and *have to*are used in different registers. It is clear from Figure 2 that in academic English, *must* occurs 590.1 times per million words, which is 2.8 times more than its frequency in spoken English (207.5 times per million words). The frequency of shall shows a similar comparison across the two registers, presenting 57.5 times per million words occurrence in academic English and only 21.1 times per million words occurrence in spoken English. Although *shall* is less frequent than *must* overall, this chart suggests that it is used more in formal or institutional discourse. This pattern aligns well with the argument that grammatical choices show a systematic register-based variation, which is especially evident in spoken and written discourse [9].

In contrast, the pattern of *have to* is opposite to the pattern of core modals (*must, shall*). The data in Figure 2 show that while it occurs 1002.2 times per million words in spoken English, the frequency in academic English is only 150.8 times per million words. *Have to* is much more frequent in spoken contexts than *must* and *shall* suggests that it is closely associated with informal interaction. Therefore, *have to* is closely related to conversational and less formal contexts.

Therefore, the rise of *have to* can be understood as a manifestation of colloquialization, since colloquialization suggests that spoken or conversational patterns are becoming increasingly crucial in general language use [1].

Table 3. Semantic and register features of selected concordance examples

Item	Example	Meaning	Style
must	We *must* act.	obligation	authoritative/ spoken
must	Parties *must* agree to delegate implementation of the agreed-upon policies.	obligation	formal/ academic
shall	The Senate *shall* appoint judges of the Supreme Court.	obligation	legal/ formal
shall	DRGs *shall* assist Regional Health Agencies in monitoring and enforcing.	obligation	institutional/ formal
have to	He is going to *have to* answer the question.	obligation	conversational/ spoken
have to	The rest of the world would *have to* adapt.	obligation	neutral/ academic

The concordance examples further support this interpretation. As shown in Table 3, both *must* and *have to* can express obligation. For example, *We must act* and *He is going to have to answer questions both convey obligation*. However, *must* often sounds stronger or more authoritative, while *have to* is more neutral and conversational. In addition, as shown in legal or institutional examples such as *The Senate shall appoint judges of the Supreme Court*, *shall* appears even more formal and institutionally marked. Apart from a diachronic frequency change, the rise of *have to* can be regarded as a register-based and pragmatic shift toward more conversational and less formal expression.

As a result, the increasing use of *have to* can be interpreted as part of colloquialization, while *must* and *shall* are largely restricted to usage in formal written discourse.

4.2. Grammaticalization and semi-modalization

Another mechanism is grammaticalization, or more specifically, semi-modalization. While colloquialization explains the register-based shift toward more spoken usage, grammaticalization helps explain why *have to* can increasingly function as an obligation marker in the modal system.

Historically, *have to* developed from a more lexical construction consisting of have plus a to-infinitive. Today, however, it has become a way of expressing obligation or necessity. In this sense, *have to* can be treated as a semi-modal or quasi-modal construction. The emergence of semi-modals

DOI: 10.54254/2753-7064/2026.HT34545

has been discussed within a grammaticalization framework, claiming that these constructions occupy an intermediate position between lexical verb constructions and central modal auxiliaries [10]. More generally, this process can be understood as a type of constructional change, in which a form that originally has a more lexical structure acquires grammatical function gradually [11].

This does not mean that *have to* has become a central modal auxiliary like *must*. There remains an important formal difference between them. Central modals such as *must* are followed by a bare infinitive, as in *must* go, while *have to* is followed by a to-infinitive, as in *have to* go. In this sense, *have to* remains between a lexical construction and a central modal auxiliary; therefore, it should be described as a semi-modal or quasi-modal.

Nevertheless, the results of this study show that *have to* increasingly overlaps with deontic *must* in function. It expresses obligation and necessity, and it is especially frequent in spoken contexts, which suggests its growing role within the modal system.

To sum up, grammaticalization should be understood here as a supplementary explanatory framework rather than as a process directly proven by the present data alone. The rising frequency and the functional overlap with deontic *must*, and strong preference for spoken contexts of *have to* are consistent with previous accounts of *have to* as a semi-modal or quasi-modal construction involved in grammaticalization.

5. Conclusion

This study has examined the diachronic change and functional relationship among *must*, *shall*, and *have to* in American English. It focused on whether core modal verbs have declined over time and whether *have to* can be understood as a partial replacement for *must* in expressing obligation or necessity.

The findings show that *shall* declines most dramatically, while *must* also shows a clear long-term decrease. In contrast, *have to* rises steadily and becomes especially frequent in spoken English. The correlation results suggest a strong association between the decline of core modals and the rise of *have to*. However, semantic coding shows that in most cases, *have to* only overlaps with deontic *must*, which is especially evident in obligation and necessity contexts. Therefore, the change should be understood as a partial functional replacement rather than a complete substitution.

This study still has some limitations. The semantic coding sample is relatively small, and the grammaticalization process of *have to* is not directly traced in detail. Future research could use a larger sample of concordance lines, compare more semi-modals such as need to and be going to, and examine how *have to* develops grammatically across different periods, registers, and varieties of English.

References

[1] Leech, G., Hundt, M., Mair, C., & Smith, N. (2009). Change in contemporary English: A grammatical study. Cambridge University Press.

[2] Leech, G. (2003). Modality on the move: The English modal auxiliaries 1961–1992. In R. Facchinetti, F. Palmer, & M. Krug (Eds.), Modality in contemporary English (pp. 223–240). Mouton de Gruyter.

[3] Love, R., & Curry, N. (2021). Recent change in modality in informal spoken British English: 1990s–2010s. English Language & Linguistics, 25(3), 537–562. https: //doi.org/10.1017/S1360674321000265

[4] Daugs, R. (2017). On the development of modals and semi-modals in American English in the 19th and 20th centuries. In T. Hiltunen, J. McVeigh, & T. Säily (Eds.), Big and rich data in English corpus linguistics: Methods and explorations (Studies in Variation, Contacts and Change in English, 19). VARIENG. https: //varieng.helsinki.fi/series/volumes/19/daugs/

[5] McEnery, T., & Hardie, A. (2012). Corpus linguistics: Method, theory and practice. Cambridge University Press.

DOI: 10.54254/2753-7064/2026.HT34545

[6] Davies, M. (2008–). The Corpus of Contemporary American English (COCA). English-Corpora.org. https://www.english-corpora.org/coca/
[7] Davies, M. (2010–). The Corpus of Historical American English (COHA). English-Corpora.org. https://www.english-corpora.org/coha/
[8] Gries, S. T. (2023). Statistics in corpus linguistics. In F. O. Unuabonah, R. O. Oladipupo, & F. O. Daniel (Eds.), Readings in corpus linguistics: A teaching and research guide for scholars in Nigeria and beyond (pp. 78–114). Kraft Books.
[9] Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). Longman grammar of spoken and written English. Longman.
[10] Krug, M. (2000). Emerging English modals: A corpus-based study of grammaticalization. Mouton de Gruyter.
[11] Traugott, E. C., & Trousdale, G. (2013). Constructionalization and constructional changes. Oxford University Press.