

**UNIVERSITY OF TARTU
DEPARTMENT OF ENGLISH STUDIES**

**PAST PARTICIPLE LEVELING IN THE PRESENT PERFECT
CONSTRUCTION: A QUALITATIVE CORPUS-BASED
ANALYSIS
MA thesis**

DARIA ZIMARIN
SUPERVISOR: *Assoc. Prof.* JANE PADRIK, PhD

**TARTU
2026**

ABSTRACT

This thesis examines past participle leveling in the present perfect construction (e.g., *I have drank*) in spoken American English. The first part of the study outlines the theoretical background, focusing on past participle variation and Construction Grammar. The second part of the thesis presents a qualitative analysis of the frequency, verb types, and discourse contexts of leveled forms in the Santa Barbara Corpus of Spoken American English. Finally, the findings are interpreted as evidence of competing constructions and stable variation, contributing to broader accounts of morphosyntactic variation. The conclusion summarizes the main results and suggests directions for future research.

Keywords: past participle leveling, present perfect, Construction Grammar, qualitative corpus analysis, constructional competition, present-day American English

TABLE OF CONTENTS

ABSTRACT	2
LIST OF ABBREVIATIONS	4
INTRODUCTION	5
1 ENGLISH VERBS, PAST PARTICIPLE LEVELING, AND CONSTRUCTION GRAMMAR.....	11
1.1 Historical English Strong Verbs	11
1.2 Synchronic Irregular Verb Paradigms in English	13
1.3 Past Participle Leveling in Present-Day English.....	15
1.3.1 Previous Studies on Past Participle Leveling in the 20 th and 21 st Centuries	15
1.3.2 <i>Have got</i> and <i>have gotten</i>	20
1.3.3 Modals and Past Participle Leveling	20
1.4 Construction Grammar and Variation.....	22
1.4.1 Construction Grammar: Theory and Terminology	22
1.4.2 Constructions, Variation, and Language Change	24
1.4.3 Intra-Speaker Variation.....	28
2 PAST PARTICIPLE LEVELING IN THE SBCSAE CORPUS	30
2.1 Corpora: SBCSAE and COCA.....	30
2.2 Method of Analysis and Data Extraction	33
2.3 Past Participle Leveling in SBCSAE	37
2.3.1 Frequency and Distribution of Past Participle Leveling in SBCSAE	37
2.3.2 <i>Have chose</i>	41
2.3.3 <i>Have went</i>	44
2.3.4 <i>Have saw</i>	47
2.3.5 <i>Have beat</i>	48
2.3.6 <i>Have spoke</i>	50
2.3.7 <i>Have took</i>	52
2.3.8 The Special Case of <i>have got</i> and <i>have gotten</i>	54
2.4 Discussion	56
CONCLUSION	68
REFERENCES.....	71
Appendix 1: Types of Recordings in SBCSAE.....	76
Appendix 2: An Example of a .cha File in SBCSAE	79
Appendix 3: List of Standard Irregular Verb Forms.....	80
Appendix 4: CLAN Commands Used to Search the SBCSAE.....	81

LIST OF ABBREVIATIONS

CxG – Construction Grammar

PP leveling – past participle leveling

SBCSAE – Santa Barbara Corpus of Spoken American English

V – verb

Vpp – Verb in past participle form

Vpt – Verb in past tense form

INTRODUCTION

As an L2 speaker of English, I often notice non-standard present perfect forms in the speech of my American friends, such as *I haven't drank* (standard: *I haven't drunk*). When I inquired whether these forms sounded natural to them, they expressed no awareness of breaking a grammatical norm. Their intuitive acceptance of such constructions sparked my interest in understanding why these patterns occur in native speech and how they fit into the broader system of English grammar.

Past-tense forms used in participle contexts have been attested for more than 500 years (Geeraert and Newman 2011: 1–2; Lass 1999: 169–170; OED 2024). Yet, the research on the phenomenon remains limited to certain contexts and varieties. Chatten et al. (2024: 4) argue that this leveling is understudied partly due to the phenomenon acquiring many distinct names throughout the years. The use of the past tense form instead of the past participle was termed participle leveling by Chatten et al. (2024: 4), Past Tense Spreading by Geeraert and Newman (2011: 1) and preterite¹ shift by Lass (1994: 89). Chatten et al.'s (2024) term seems the most appropriate as it does not signal the direction of the shift, since thus far, there is no conclusive evidence on that. Nevertheless, the term *participle leveling* proposed by Chatten et al. (2024) can still be extended, as it does not specify the type of participle, although there are at least three kinds of participles in English. Inspired by Chatten et al. (2024), a more specific term is used throughout this thesis – past participle leveling (henceforth, PP leveling).

There are two main, larger corpus-based studies that aim to quantify PP leveling. One study (Geeraert and Newman 2011) uses COCA and BNC, which consist mainly of written texts with smaller spoken parts. Additionally, Geeraert and Newman (2011) utilize the written texts on the web as a corpus. Another, newer study (Chatten et al 2024) fills some gaps left by Geeraert and Newman (2011) and describes both external and internal contexts and influences

¹ Preterite is another name for the past simple form.

on PP leveling using the British corpus DECTE, which consists of Tyneside English sociolinguistic interviews, the American Switchboard with phone conversations, and the American PNC (Philadelphia Neighborhood Corpus) with sociolinguistic interviews. The corpora in both studies have certain limitations. The BNC and COCA consist mostly of written text, with the spoken parts smaller and drawn mostly from TV and radio recordings, which may represent less spontaneous and less vernacular speech. Even though the corpora in Chatten et al. (2024) are certainly more vernacular than those in Geeraert and Newman (2011), the conversations in both sociolinguistic interviews and phone conversations were elicited (i.e., the speakers were asked specific questions or given topics to discuss), and elicited speech may be considered less spontaneous. While both studies provide a quantitative overview of PP leveling in British and American Englishes, they lack detail in their examination of the meaning and forms of PP leveling compared to those of a standard construction. Therefore, an investigation of a spoken corpus of spontaneous speech may provide more insight into PP leveling.

The Santa Barbara Corpus of Spoken American English (henceforth SBCSAE) is used in this study as it consists mostly of spontaneous American speech. Although corpora of spoken language are not new in linguistics research, most spoken corpora are drawn from speech produced in more careful, possibly scripted or prompted contexts. The most well-known and widely used corpora in English linguistics, BNC and COCA, also include spoken sections that constitute approximately 11% and 13% of the total corpus, respectively, and consist of TV, radio, and movie speech (Davies 2008). Even though the creator of BNC explicitly states that they estimate 97-98% of speech to be unscripted (Davies 2008), the actual ‘unscriptedness’ of the speech is difficult to prove or check. Davies (2008) also points out that in the BNC, the main priority was a larger size, even if that meant some speech in the corpus would be scripted rather than spontaneous. Creating a spoken corpus of spontaneous speech requires finding the subjects (and obtaining their informed consent), recording their speech in real-life situations,

transcribing, and converting the material into a machine-readable format – a quite laborious undertaking. This is the dilemma creators of spoken corpora often face: either collecting unscripted speech while keeping the corpus smaller, or compiling a larger corpus, sacrificing control over spontaneity. The Santa Barbara Corpus of Spoken American English follows the former principle of collecting unscripted speech.

The SBCSAE, which mostly comprises recordings of naturally occurring, spontaneous speech, is used in this study to investigate the use of past participles in the present perfect construction in spoken American English. Due to the smaller size of the corpus (249,000 words, according to Du Bois et al. (2000–2005)), a qualitative approach was employed. American English was chosen as the focus of this study for two reasons. Firstly, the phenomenon was initially observed by the author in the American variety. And secondly, American English tends to be more leveled, i.e., have less regional/dialectal variation, compared to British English. Even though the SBCSAE corpus is significantly smaller compared to COCA or BNC, the speech in it is more spontaneous and informal, including recordings of conversations between friends, family arguments, task-related talks, and more.

Construction Grammar (often abbreviated CxG) was chosen as the theoretical framework since it allows for a qualitative investigation into the meaning and use of grammatical constructions. In this framework, linguistic knowledge is understood as a network of constructions, defined as pairings of form and meaning (Goldberg 1995: 1; Croft and Cruse 2004: 257; Hilpert 2013: 5). CxG distinguishes itself from both generative and other usage-based frameworks by examining formal, semantic, and pragmatic features of constructions, rather than focusing on structure or function alone (Goldberg 2013: 11–13). The main assumption of the framework is the absence of strict division between lexis and grammar: lexical items and grammatical patterns are both treated as constructions and understood as differing only in their degree of schematicity and complexity (Goldberg 1995: 7). Such a

perspective makes Construction Grammar particularly suitable for the analysis of morphosyntactic variation, such as PP leveling, where alternative forms may express similar meanings but differ in structure and contextual distribution. More recent studies have further shown that constructions may exhibit variation conditioned by usage, frequency, and context. Moreover, constructions can be characterized as abstract, ‘lossy’ mental representations within a high-dimensional conceptual space (Goldberg 2019: 7; Hilpert 2025: 150–151). Because CxG treats variation as an inherent property of constructions and foregrounds meaning in use, it provides a well-motivated framework for qualitative analysis, allowing close examination of semantic properties, pragmatic function, and discourse contexts, as well as competition between constructional variants.

The application of the CxG framework has so far been limited at the Department of English Studies at the University of Tartu. One earlier MA thesis that utilizes the CxG framework is Roomäe’s (2021) thesis, *How Come That’s a Construction? A Construction Grammar Approach to the How Come Construction in Spoken American English*. Roomäe (2021) uses a corpus-based CxG approach to analyze the *how come* construction in spoken American English, focusing on its form, meaning, and distribution. While the present study similarly adopts a CxG framework to examine a spoken construction, Roomäe employs a quantitative approach and focuses on internal and external form, the former addressing the construction’s formal and semantic properties out of context, and the latter examining the surrounding context (e.g, surrounding elements, register, discourse function, etc.). In contrast, the present study focuses on the qualitative analysis of each example in the corpus and on the discussion of variation and competing constructions.

Examining PP leveling may be especially useful for furthering the understanding of several aspects of Construction Grammar. Firstly, such an investigation may provide insights into the stability and flexibility of verbal morphology, i.e., how flexible the form of the present

perfect construction is in native speakers' minds. Examining both intra-speaker and inter-speaker variation can provide such insights, prompting the discussion of how speakers mentally represent constructions and their associated variants. Secondly, tackling the standard present perfect and PP leveling as competing constructions may allow us to make predictions about how language change occurs, as well as to describe the reasons for the persistence of non-standard forms. Therefore, the investigation into PP leveling allows for exemplifying and advancing the discussion of central issues in CxG, such as constructional competition, frequency, entrenchment, and variation.

The main aim of this thesis is to further the understanding of the PP leveling by examining it qualitatively, through the CxG framework, in spontaneous speech. A qualitative analysis can bring valuable insights that a quantitative analysis cannot, such as insight into intra-speaker variation. Additionally, this study aims to advance the discussion of constructional variation and constructional competition by examining the case of PP leveling. To do that, this study poses the following research question: What are the meaning, usage contexts, and formal properties of the PP leveling construction in spoken American English compared to the standard construction? The research question can be further divided into three aspects: meaning, context, and form of the leveled construction. The meaning of the construction will be examined by comparing it to the standard construction, where possible, and by focusing on the semantics of tense and aspect in the construction. The context will be discussed and exemplified by examining the linguistic elements surrounding the construction (e.g., neighboring or overlapping constructions) and the conversational style/register. The form will be analyzed by creating a constructional schematic representation and by discussing the formal differences between the PP-leveled and standard constructions.

Based on the research question, it is expected that PP leveling will be rare but systematic in the SBCSAE, i.e., certain verb lemmas and patterns will appear across PP-leveled

constructions. Moreover, certain verbs and verb classes may exhibit a tendency to level more often than others. Additionally, PP leveling is expected to occur more frequently in casual-style conversations. Syntactic context may also influence the probability of PP leveling: based on the findings of previous studies (Chatten et al 2024; Geeraert and Newman 2011: 15–16), modals and first-person subjects are expected to co-occur with PP-leveled constructions. Finally, the possibility of competition between the two constructions and its possible effects on further variation and change will also be discussed.

The first chapter of this thesis provides a description of the English verb system, both from a diachronic and synchronic perspective, which will serve as background for the examination of variation over time and for discussing the possible influence of PP leveling on the English verb system later in the thesis. Moreover, this chapter discusses the CxG framework, defining its key terminology and tying it in with variation and language change. The second chapter presents the empirical analysis. The analysis is qualitative and based on the CxG principles discussed in the first chapter. In the second chapter, the corpora and the procedures for data collection and analysis are described in detail. This is followed by a summary of the frequency and distribution of the PP leveling. The examples from the SBCSAE corpus are then presented and analyzed individually. The discussion of examples is organized according to the verb appearing in the construction. The concluding section of the chapter, Discussion, presents the results and their interpretation in a summarizing manner to answer the research question.

1 ENGLISH VERBS, PAST PARTICIPLE LEVELING, AND CONSTRUCTION GRAMMAR

Anderson (2001: 228) states that “past changes can commonly be found to be reflected in synchronic alternations”, meaning that synchronic variation and historical language change can be connected. Therefore, this section provides an overview of English verb patterns throughout history and discusses verb paradigms and their leveling patterns, as these topics are relevant to PP leveling. Moreover, since this thesis employs Construction Grammar to analyze PP leveling, this chapter provides an overview of key terms and theories within the framework.

The first section of the chapter discusses historical strong verbs and their possible implications for PP leveling. The second section discusses verb paradigms and leveling patterns found in present-day English. The third section focuses on studies that specifically discuss PP leveling, including *have got* and *have gotten*, as well as the co-occurrence of modals with PP leveling. The chapter concludes with a CxG section that gives an overview of the CxG framework and defines key terminology. It also discusses how variation and language change fit into CxG.

1.1 Historical English Strong Verbs

The use of the past tense form rather than the past participle may be linked to English strong² verb paradigms and the history of their formation. English grammar textbooks generally classify verbs into two classes: weak (regular) and strong (irregular) (e.g., Quirk et al. 1985: 96–114). This classification is based on how verbs form the past tense and the past participle. Weak verbs are the majority of verbs in English and form both past tense and past participle by adding the suffix *-ed* (e.g., *walk-walked-walked*), while strong verbs exhibit stem alternations (e.g., *drink-drank-drunk*) (Taylor 1994: 1; Anderwald 2009: 51; Anderwald 2012: 258). The

² ‘Strong verbs’ is used in historical contexts and ‘irregular verbs’ in synchronic contexts.

strong verbs are particularly of interest here, as they are descendants of an Indo-European ablaut system, while weak verbs are a Germanic feature. It must also be mentioned that some verbs in English have normative³ weak and strong conjugation. According to Taylor (1994: 1), verbs such as *lie-lay*, *lay-laid*, *hung/hanged*, and *shone/shined* may conjugate in both paradigms, possibly due to lexical confusion (e.g., *hung* and *hanged* have different meanings). Other verbs, such as *crept/creeped*, *wept/weeped*, that have two ways of conjugating and no semantic difference between the two forms, are termed semi-weak or irregular weak (Taylor 1994: 1; Lass 1999: 172–175). This thesis is particularly concerned with strong verbs, as PP leveling can be studied only with those verbs that have three distinct forms in their conjugation paradigm.

In English, strong verbs form their past tense and past participle differently for various historical reasons. Among them, one crucial influence on English verb paradigms can be pointed out – the disappearance of Old English inflectional forms. According to Lass (1994: 81), standardization in English began in the 14th century and was largely completed by the end of the 18th century. However, the process of this standardization is somewhat obscure – it is unclear exactly how it played out. Examining the historical strong verb paradigms, it is evident that after the disappearance of Old English inflections, certain changes had to occur in how strong verbs formed their past tense and past participle, as English began to prefer invariant-base morphology, i.e., morphological paradigms that do not exhibit stem alterations (e.g., invariant-base: *play-played*, stem alteration: *drink-drank*). In addition, the grammatical distinction of number disappeared, meaning that verbs no longer needed to agree with nouns in singular or plural forms. Instead, tense became the primary feature expressed through verb inflection. As a result, the verb paradigms underwent simplification. The simplification process included filtering out strong verb forms, i.e., narrowing the number of forms strong verbs should

³ ‘Normative’/‘non-normative’ and ‘standard’/‘non-standard’ are used interchangeably in this thesis.

have, as well as selecting the Old English forms to be preserved. Such processes involved variation, including, for example, a shift of the preterite to the past participle (e.g., *I have sang*).

While most changes in verb paradigms are attributed to the overall changes of the inflectional system that happened at some point between Old English and Modern English, in some cases, verb paradigms changed due to non-grammatical reasons. The most intriguing case of historical PP leveling is possibly the use of the form *drank* as a past participle. What is notable here is that several sources ascribe this shift to ‘taboo avoidance’. The form *drunk* may have been avoided due to its association with the state of intoxication by alcohol (Atwood 1953: 43; Lass 1999: 170). While Lass places such use between the 15th and 19th centuries, The Oxford English Dictionary includes the past participle *drank* as a non-standard or regional form that has been in use since the 1700s and remains in use today (OED 2024). This may mean that the verb *drink* is more accepted in the perfect construction in its leveled form.

Most *write-wrote-written* paradigm and *drink-drank-drunk* paradigm verbs have exhibited PP leveling throughout the centuries, beginning in Middle English. According to both Lass (1999: 169–170) and the OED (2024), the use of the *a* instead of the *u* in the stem is attested both diachronically and synchronically, although mostly in dialectal or non-standard use in the latter. Type *write-wrote-written* verbs exhibit similar behavior. The OED lists the form *wrote* as a past participle in use since the 1500s, with a note that it is “now regional or nonstandard” (OED 2024). Such historical records suggest that PP leveling has existed for centuries, originating from linguistic variation that predated the standardization of English. This variation in verb forms may have been more widespread than today, as today’s prescriptive norms discourage them.

1.2 Synchronic Irregular Verb Paradigms in English

Examining contemporary English verb paradigms reveals intriguing patterns related to PP leveling. There are different classifications of irregular verbs according to their past tense

and past participle formation paradigms; however, for the purposes of this thesis, the classification proposed by Anderwald (2009), adapted by Geeraert and Newman (2011), is used (Figure 1). Figure 1 demonstrates not only all existing English irregular verb paradigms, but also the percentage of English irregular verbs falling into these categories. According to Anderwald (2009) and Geeraert and Newman (2011), class (b) tends to attract verbs from other classes through leveling – a pattern also observed in PP leveling, where a shift from class (e) or class (d) to class (b) happens (although, a rarer shift from class (c) to (a) is also possible). Both studies mention several reasons for the attractiveness of class (b), mainly its size and the historical behavior of Germanic strong verbs (see section 1.1).

(a)	infinitive 'cut'	simple past 'cut'	past participle 'cut'	14.4%
(b)	infinitive 'meet'	simple past 'met'	past participle 'met'	48.5%
(c)	infinitive 'beat'	simple past 'beat'	past participle 'beaten'	0.6%
(d)	infinitive 'come'	simple past 'came'	past participle 'come'	1.2%
(e)	infinitive 'drink'	simple past 'drank'	past participle 'drunk'	35.3%
				100%

Figure 1. Irregular verb paradigms with the percentage of English irregular verbs falling into each class (Geeraert and Newman 2011).

The Carstairs' (1983) paradigm economy principle is also worth mentioning, as this shift from class (e) or class (d) to class (b) may indicate that such leveling is underway. The

principle of paradigm economy states that although a language's paradigm inventory can be enormously large in theory, languages tend to employ the number of paradigms that is close to a logical minimum. To determine the maximum possible number of paradigms for a given word class, one has to list every possible morphosyntactic slot (e.g., each case-number combination) and count how many different suffixes (or forms) are available for that slot. Since the verbs found in the PP leveling most often change their class from (e) to (b), it may indicate language change driven by the paradigm economy principle. The existence of PP leveling may then be explained by the inflectional system of English striving to reduce the amount of irregular verb paradigms.

1.3 Past Participle Leveling in Present-Day English

This subchapter provides an overview of PP leveling studies focusing on the phenomenon in the 20th and 21st centuries. The first section provides a general overview of the studies on PP leveling. The second section discusses the case of *have got* and *have gotten* as a shift that has already happened. Finally, the third section discusses the co-occurrence of PP leveling and modals found in previous studies.

1.3.1 Previous Studies on Past Participle Leveling in the 20th and 21st Centuries

PP leveling is a relatively understudied phenomenon, although it has existed for centuries (Geeraert and Newman 2011: 1–2). Modern English investigations of this phenomenon prior to Geeraert and Newman's (2011) study are mostly credited to Lass (1999: 166–175; 1994: 89–90) and Anderwald (2012: 257–260). In his investigations, Lass admits that variation in past form and past participle usage is found in both Middle and Modern English. In his example (Table 1 below), all existing paradigm variation is demonstrated using the verb *sing*, which is a *drink-drunk-drunk* paradigm verb. The 'Historical' paradigm is described as the one typical of Old English strong verbs. The 'Western' variation is the one developed

through leveling, where the past form takes the past participle *u*, which originated in the West of England in the fourteenth century as a rarely used form and gained prominence later, towards the end of the Early Modern English period (Lass 1994: 89). Both Lass (1994: 89) and Anderwald (2012: 260) argue that the preterite shift paradigm, the one under investigation in this thesis, and the ‘crossover’ paradigm originated in Middle English as a result of the loss of the systematic inflection-based verb paradigms of Old English and the subsequent uncertainty among speakers. With this, the authors demonstrate that a wide range of variation, rooted in earlier periods of English, exists in present-day English. This kind of variation and uncertainty caused by changes in verb paradigms between Old and Middle English may be one of the reasons for the existence of PP leveling today.

Table 1. Verb form variation in Early Modern English (Lass 1994: 89)

past form	past participle	type of paradigm
<i>sang</i>	<i>sung</i>	historical
<i>sung</i>	<i>sung</i>	‘Western’
<i>sang</i>	<i>sang</i>	preterite shift
<i>sung</i>	<i>sang</i>	‘crossover’

Unlike PP leveling, its reverse counterpart, the ‘crossover’ paradigm, as termed by Lass (1994), is much more widely studied in both British and American English (e.g., Anderwald 2011; Anderwald 2009: 63; Bybee and Moder 1983; Bybee 1985; Bybee 1995). Chatten et al. (2024: 4) propose two reasons for that: the multiplicity of names for the phenomenon (see the Introduction) and its rarity. The two most recent corpus-based studies on PP leveling are Chatten et al. (2024) and Geeraert and Newman (2011), both of which use larger corpora consisting of speech with different levels of spontaneity.

Geeraert and Newman (2011) investigate PP leveling across two corpora, the BNC and the COCA, and additionally use the Google Search Engine to find examples of the phenomenon on the web. The authors investigate PP leveling not only in the present perfect but also in the past perfect construction. The results of their study clearly show that PP leveling is most frequent in the spoken part of the corpora. This result might be expected, as PP leveling seems to be a mostly non-standard phenomenon; thus, it is more likely to appear in spoken language. Nevertheless, Geeraert and Newman's study has several limitations. The most prominent of them is the method used to identify PP leveling in the corpora. The authors used combinations of word forms or 'fixed expressions' to identify PP leveling in the corpus, e.g., *have rode, has rode, had rode* (Geeraert and Newman 2011: 4). As the authors admit, this simplification of the search method limits the results by several factors. Firstly, it does not consider the cases where the auxiliary verb is separated from the lexical verb by other tokens. Secondly, it does not include any negative forms that may occur as PP leveling, e.g., *I haven't drank*. Moreover, it is highly unlikely that the authors could predict all the possible verbs that may occur in PP leveling constructions. All in all, this methodological approach may have limited the number of PP leveling instances found in the corpora, and with a more thorough search, this number could have been significantly larger, encompassing a broader range of verbs.

Chatten et al. (2024) build on Geeraert and Newman's (2011) and other earlier studies and attempt to test the previous findings/hypotheses across a relatively more diverse set of corpora in both British and American varieties. The main conclusion of the study is that the variation in the perfect construction is largely systemic and that any verb can vary. As in other studies, the authors report that lower-frequency verbs tend to exhibit leveling more often. Moreover, it was found that verbs with greater phonological similarity between the past tense and the past participle tend to level more. For example, the verb paradigm *write-wrote-written* would level less than the verb paradigm *drink-drank-drunk*. Modals with non-finite perfect

(more on that in section 1.3.3) were found to induce leveling. Subject-person dependency was found only in one of the two US corpora. In Switchboard, third-person subjects co-occurred with leveling less than first-person subjects. Intervening negation was found to induce PP leveling only in Switchboard, whereas the other two corpora showed an opposite pattern. Additionally, the authors report intra-speaker variation, in which one speaker uses both the standard and the construction with the leveled past participle in their speech. Intra-speaker variation was observed in all three corpora of the study and across multiple speakers, although only three examples were provided. These findings showed that PP leveling is a complex phenomenon influenced by many linguistic factors. Nevertheless, there are some limitations also in Chatten et al.'s (2024) approach.

Chatten et al. (2024) used two spoken American English corpora, PNC (Philadelphia Neighborhood Corpus) and Switchboard, as well as DECTE for the British variety. Even though the corpora are certainly more vernacular than Geeraert and Newman's (2011), the conversations in the corpora are still elicited, which may indicate a lesser degree of spontaneity. PNC consists of sociolinguistic interviews, and Switchboard consists of phone conversations with topics assigned by the researchers. The methodology used to extract data from the corpora is also lacking in certain aspects. The perfect constructions were extracted using a Python script that searched for combinations of the verb *have* and verbs from the authors' irregular verb list. Some verbs may have been excluded because they were not on the list. Moreover, it was defined in the search that there should be no more than one word between *have* and the past participle. This may have prevented the script from outputting some instances of the perfect, where there are several words between *have* and the past participle, which is especially plausible in spoken, informal contexts (e.g., *I have never ever said that*).

There are several older sociolinguistic investigations of non-standard irregular forms that deal with specific dialects of English. These studies examine the influence of social factors,

such as age and gender, on the use of non-standard forms. Many studies, such as Wolfram and Christian's (1975) Appalachian English study, Feagin's (1979) Anniston English study, and Shnukal's (1978) Cessnock study, found that social factors do not influence the use of non-standard irregular forms. Both Eiskovits's (1987) and Peters' (2009) studies of Australian and New Zealand English, however, conclude that there is some influence on the non-standard use of irregular forms by age, gender, style, and context. Eiskovits (1987) found that female speakers use fewer non-standard irregular forms in formal contexts than male speakers. However, the corpus in Eiskovits's study consisted of only 40 adolescents, so generalizations based on such a small number of speakers should be approached with caution. A more recent study by Peters (2009) also indicates there is a tendency for younger speakers to use more non-standard forms in New Zealand and Australian Englishes. It may, thus, be that social factors do influence non-standard verb usage in the Southern hemisphere. Nevertheless, there seems to be no consensus on whether sociolinguistic factors have an influence on standard/non-standard preference in other varieties of English.

A regional survey by Atwood (1953) aimed to provide a comprehensive overview of non-standard verb forms across different regions of the US. A relevant finding of this study is the lack of clear regional variation borders. The study found a tendency to preserve older verb forms in regions such as the New York City area, eastern Pennsylvania, and more urbanized parts of New England, but, beyond that, variation does not seem to be strictly defined by regional borders. The study reports that some forms, such as *drove*, *shrunk*, *drank*, and *drunk*, are used as both past simple and past participle forms throughout the US (Atwood 1953: 43). Moreover, the author reports leveling along both PP leveling and crossover paradigms (see Table 1) for such verbs as *swim*, *ring*, *see*, *do*, *ride*, *wear*, *write* and *break*. All in all, this survey showed that although variation is present across several regions, no area associated with the variation could be identified, and it appears to occur across all of the US.

1.3.2 *Have got* and *have gotten*

Even though there is debate on the origins of *have gotten* and the reasons for its retention or revival in American English, it is clear that this form is more often found in American English and is almost completely erased from British English (Anderwald 2020: 193; Hundt 2009: 20). *Have gotten* is now widely recognized as an Americanism, with *have got* also being used in North America, but the two constructions having distinct meanings (Fowler and Burchfield 1998: 338; Hundt 2009: 20). In present-day American English, *have gotten* is usually used with the dynamic meaning (i.e., ‘obtained’), while *have got* with the stative one (i.e., ‘possess’). Nevertheless, this distinction is not clear-cut, as this construction may be challenging to study due to *have got* having many distinct uses and meanings (Fowler and Burchfield 1998: 338). Regardless, an attempt was made to categorize *have got* and *have gotten* by meaning in this study.

Geeraert and Newman (2011: 5) report that *got* was used as a past participle in the BNC corpus 99% of the time, which is why the authors treated *got* as a standard past participle. However, it is important to note that BNC is a corpus of British English, in which it has already been established that *got* as a past participle has become standard. It is, however, not the case with American English, which still uses *gotten*, as established in previous studies. Even though Geeraert and Newman decided to exclude *have got* from their PP leveling inquiry, it seems necessary to examine this phenomenon in American English. Similarly to Geeraert and Newman’s study, *have got* was not considered a PP leveling case in this thesis; instead, the occurrences and frequency of *have got* were reported separately.

1.3.3 Modals and Past Participle Leveling

Previous studies agree that modals influence PP leveling. An earlier study of Sydney English by Edina Eiskovits (1987), the study of COCA and BNC by Geeraert and Newman (2011), and Chatten et al.’s (2024) comparison of British and American varieties all found a

correlation between modal auxiliaries and PP leveling. Eiskovits' (1987) study dealt with Australian adolescents and found that 20% of the modal-perfective constructions (e.g., *would have gone*) used the past simple rather than the past participle form of the lexical verb. Compared to the non-modal perfective structure, with a PP leveling percentage of 12.2%, there is a difference in frequency of almost 8% (Eiskovits 1987: 23). Even though Eiskovits's study may be less relevant, as it is older and deals with Australian English, the findings suggest that modals do influence PP leveling, though the impact is relatively modest.

According to Geeraert and Newman's (2011: 15–16) findings, the patterns in COCA and BNC are similar. However, their percentages were much higher than those of Eiskovits (1987). For example, the authors report that 82.7% of all *have went* cases in COCA included a modal auxiliary in the construction. Compared to the 40.9% of *have gone* co-occurring with modals, it seems that this distribution needs to be studied further. Geeraert and Newman (2011) report that all the corpora used in their study (BNC, COCA, and web texts) exhibited similar patterns. It is unclear whether COCA showed the most dramatic difference, as the authors provide no statistics for the other corpora. Nevertheless, this study's findings suggest that the modals may influence PP leveling.

Chatten et al. (2024) report the co-occurrence of PP leveling and modals as their main finding regarding internal context. They found that among the cases where *have* was non-finite in the PP leveling construction, 94% of instances were preceded by a modal (Chatten et al 2024: 26). They also explicitly state that the reason for that may be due to “modal + *have* perfect” having a different syntactic structure (Chatten et al 2024: 11). The authors go even further to suggest that the combination of a modal and a phonetically reduced auxiliary *have* in the modal perfect (e.g., *woulda*, *shoulda*, *coulda*) can be reanalyzed as an adverb. In that case, the modal+auxiliary unit would not take on any finite morphology, and the verb to be tensed would be the main verb (see Figure 2). This reanalysis justifies the use of the past tense in the

construction. However, the authors admit that modals alone cannot adequately explain the leveling, as not all cases of PP leveling include a modal.

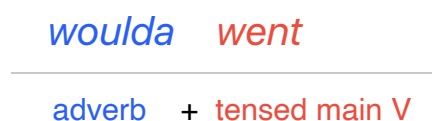


Figure 2. The reanalyzed construction of the modal perfect.

1.4 Construction Grammar and Variation

This subchapter focuses on the CxG framework. The first section gives a more general introduction to the theory and terminology. The second section describes how the framework treats variation and language change. The final, third section, deals with intra-speaker variation, with a focus on a morphosyntactic angle.

1.4.1 Construction Grammar: Theory and Terminology

Construction Grammar (abbreviated as CxG) appears to refer to a number of related approaches within usage-based theories in linguistics. Since CxG as a term encompasses more than one approach, Goldberg (2013: 12) proposed referring to Construction Grammar as ‘constructionist approaches’. In her opinion, this allows us to move away from extensive labelling and encourages collaboration between researchers from related fields. However, this thesis will employ the term ‘Construction Grammar’, as it is more recognizable and widely used across sources.

CxG is an approach that distinguishes itself from both the generative and usage-based approaches by its main principle: the emphasis on the pairing of form and meaning. Generative approaches tend to de-emphasize function, and usage-based approaches tend to de-emphasize form (Goldberg 2013: 11), while CxG attempts to incorporate both. It is also important to note that CxG includes both lexical and grammatical constructions and does not assume any divisions between them. Even though syntactic and lexical constructions may vary in their

complexity, they are still regarded as pairings of form and meaning that share basic common principles (Goldberg 1995: 7). As such, CxG is a quite universal approach, applicable even to morphosyntax.

The idea of a ‘construction’ is central to CxG. As already mentioned, in CxG, grammatical constructions are defined as pairings of form and meaning, and they may consist of words, morphemes, and larger phrasal units (Goldberg 1995: 1; Croft and Cruse 2004: 257; Hilpert 2013: 5; Iwata 2014: 650). The schematic structure of a grammatical construction is demonstrated in Figure 3 (Croft and Cruse 2004: 258). The ‘meaning’ refers to everything that a construction (a fixed way of saying something) usually does or expresses. This may allude to both the context and the features of discourse, in which the construction is found (Croft and Cruse 2004: 258; Hilpert 2013: 6). This means that to investigate constructions, one must also investigate their internal meanings, broader discourse contexts, and surrounding linguistic elements. This thesis focuses on the syntactic and semantic properties, as well as the conversational context and register of both the standard and PP leveling constructions.

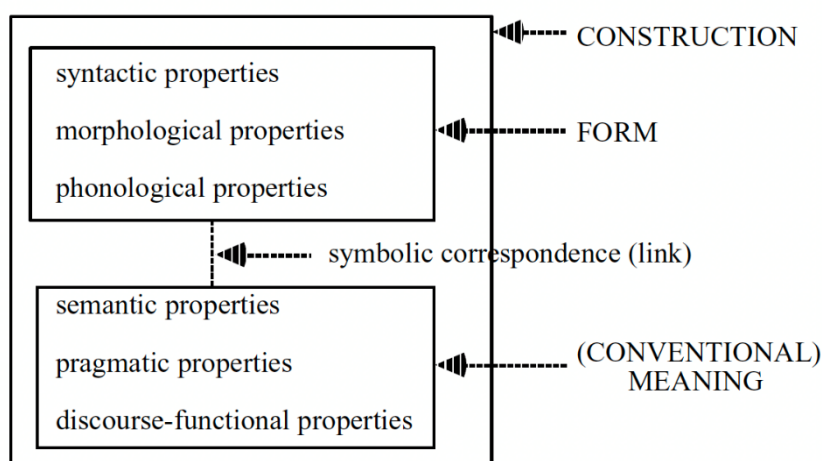


Figure 3. The structure of a construction (Croft and Cruse 2004: 258).

More recently, Goldberg (2019) has proposed a revised definition worth discussing, as it has implications for the roles of frequency and variation in CxG. “[C]onstructions are understood to be emergent clusters of lossy memory traces that are aligned within our high-

(hyper!) dimensional conceptual space on the basis of shared form, function, and contextual dimensions” (Goldberg 2019: 7). Hilpert (2025: 150–151) provides a particularly clear explanation for this definition. ‘Lossy’ memory here implies that the constructions stored in the speaker’s mind are more abstract than the actual representation of it in language, i.e., the utterance produced with the construction. Hilpert (2025: 150) also argues that Goldberg insinuates variation with ‘high-dimensional conceptual space’. It is implied that constructions vary in form and meaning; thus, knowing a construction means knowing its variations. Moreover, the definition mentions context, indicating that it is relevant to variation in constructions. As Hilpert (2025: 151) put it: “The constructions that a speaker is likely to encounter in one context may be near-absent in another, depending, for example, on the formality of the situation or the social distance between the communicators”. This idea suggests a dependency between the discourse situation or contexts and variation.

Constructions do not exist separately; they form larger units. A combination of constructions that forms a larger/longer utterance was termed a construct or construct-icon by Hoffman (Hoffman and Trousdale 2011: 2; Hoffman and Turner 2025: 1). It is argued that to know a language, one must have a mental relational network of constructions, as well as understand how they combine to form constructs. Thus, a construct is also a mental representation of the way constructions can combine. Moreover, Hoffman and Turner (2025: 2–3) point out that most often, speakers must combine multiple constructions to form a construct, but there are rare cases of one-construction constructs, such as “Thank you!”. Therefore, context is central in discussing constructions, as it may help identify larger constructs at play.

1.4.2 Constructions, Variation, and Language Change

Variation in Construction Grammar means that there are two or more variables within a construction, exemplified by PP leveling (i.e., there are two options for the present perfect

construction: have + Vpp and have + Vpt). Hoffman and Trousdale (2011: 7) argue that the choice between the two alternative variables can be influenced by independent factors. The notion of inherent variability describes the inherent feature of any language to express the same meaning in multiple alternative ways (Hudson 2007: 383; Hoffman and Trousdale 2011: 3). Hudson (2007: 384) argues that inherent variability should be studied in three dimensions of linguistic analysis: structural – the two alternatives can express the same meaning but have different structures, contextual – the two alternatives can appear in different social or linguistic contexts, and statistical – the two alternatives have different frequency or probability. Thus, it seems highly unlikely that any two variants can be completely equal in respect to all those dimensions.

It has been established that variation always exists on the level of constructions – there are often more than one form that can express the same meaning. However, there is a need to establish how these multiple forms can coexist in a language. Goldberg (2019) has proposed treating synonymous constructions as competitors. The idea behind competing constructions is that constructions expressing the same (or similar) meaning-in-context are always competing in the speaker's mind. The choice between the two constructions is made due to statistical preemption – speakers hear the conventional form more frequently, so it dominates over the novel form that feels unnatural to the speaker (Goldberg 2019: 74–76). In connection with variation and language change, competition may be viewed as a mechanism that selects among alternative constructions over time. Nevertheless, it should be noted that statistical preemption is probabilistic, i.e., it does not determine which form will be used in competition. In real language use, one form may dominate due to its frequency, while the other may still survive in slightly different pragmatic or social contexts.

Another central concept in variationist-constructionist approaches that aim to explain the coexistence of two variants of the same construction in language is the often-quoted and

widely accepted Goldberg’s (1995: 67) Principle of No Synonymy, which states: “if two constructions are syntactically distinct, they must be semantically or pragmatically distinct.” However, there is both supporting and opposing evidence for the principle. For example, Uhrig (2015) argues that treating the No Synonymy Principle as a dogma can be misleading. With such examples as dative alternation (e.g., *give John the letter* vs. *give a letter to John*) and particle placement (e.g., *give up the ship* vs. *give the ship up*), Uhrig (2015) demonstrates that meaning differences are often neutralized in context and in these cases, speakers’ choices are influenced by processing, frequency, or priming rather than semantic differences within constructions. In contrast, Klavan (2021) finds evidence in support of this principle with her investigation of the Estonian exterior locative case, which demonstrates that the formal alternations of the locative case depend not only on the Landmark phrase (e.g., in *laua peal* (“on the table”), *laua* (“table”) is the Landmark phrase), but also on stylistic, situational and register factors. Klavan (2021) and Uhrig (2015) agree on one thing: frequency and ease of processing influence the choice of an alternation. Therefore, it seems that the Principle of No Synonymy should not be taken as an inherent property of all constructions, but rather an observed tendency.

As we have seen in the definitions of ‘construction’ and the theory of variation, frequency plays a central role in the study of variation in constructions. However, high frequency seems to be associated with both driving language change and having a conserving effect (conservatism via entrenchment, as in Goldberg 2019). Bybee (2006b: 728, 2006a: 271-272) reports that morphosyntactic constructions are more resistant to change when they have higher frequencies. Both Bybee (2006a: 728) and Hilpert (2025: 153) argue that irregular verbs of high frequency resist regularization. Hilpert (2025: 153) argues that verbs with higher frequency, such as *keep-kept*, retain their irregular past, while lower-frequency verbs, such as *leap-leaped*, adopt the regular past. Higher-frequency tokens are more entrenched in speakers’

minds; thus, speakers have a clearer representation of their distribution (Hilpert 2025: 154; Goldberg 2019: 54). However, high frequency can also contribute to the non-predictability of constructions. One example provided by Hilpert (2025: 154) is the construction *I don't know* that due to its high frequency, often appears as *dunno*. There seems to be no conclusive evidence as to why certain high-frequency constructions exhibit a conserving effect, while others exhibit the opposite tendencies.

Variationist linguistics has established that language change is always accompanied by variation; however, variation does not always point to language change (Gardiner and Nagy 2017: 78; Chambers 2013: 297–298). Gardiner and Nagy (2017: 78) report that at times, natural variation persists over time but does not facilitate language change. The authors report on Wallenberg's theory, which states that competing forms are necessary for variation to exist. Within this theory, variation can lead to either language change, in which one form becomes the standard, or specialization, in which both forms are used in different contexts or with different meanings. According to this theory, stable variation is rare, as languages prefer forms with distinct meanings. Moreover, according to the Continuity Hypothesis, the variation can be stable only if it is influenced by at least one continuous factor. For example, Campbell-Kibler (2007) examines the *-ing* and *-in* variation and demonstrates that the (ING) variable is influenced by a range of gradient and context-dependent social meanings rather than a simple standard/non-standard binary variable. These continuous conditioning factors explain why the variants continue to coexist, rather than specialize and replace each other. In contrast, if a variable is influenced only by categorical factors, such as sex (i.e., male/female), it most likely points to language change.

Nevertheless, modeling some factors as categorical/continuous can be problematic. For example, as Gardiner and Nagy (2017: 82–83) admit, style is a category that can be modeled as continuous or categorical, depending on how the variable is analyzed. Some of the earlier

studies used binary distinctions, such as careful/casual and narrative/casual (e.g., Labov 1966, 1984, 2001). However, Schilling (2013: 327–338) argues that style is dynamic, gradient, and interactionally dependent, modeling it as continuous. Nevertheless, there is currently no consensus on whether style should be treated as a continuous factor. Due to the nature of the corpus used in this study and how the variable is analyzed, it was decided to model style as a continuous factor, from most casual to most careful (see the modeling of style in section 2.2).

1.4.3 Intra-Speaker Variation

Intra-speaker variation is defined as variation within a single individual’s linguistic behavior. Intra-speaker variation is understood as a natural part of language. This phenomenon has been approached through different linguistic frameworks, including phonetics, morphosyntax, and sociolinguistics, among others. This variety of approaches has demonstrated that intra-speaker variation is a complex phenomenon that can be interpreted in multiple ways (Ulbrich and Werth 2021: 9–12). Many argue that intra-speaker variation is not a random ‘performance error’ but a structured phenomenon that can be ideological, grammatical, cognitive, or interactional, depending on the analytical framework (Hernández-Campoy and Cutillas-Espinosa 2013: 81–97; Cognola et al 2019: 2–3, 16–18; Bülow et al 2021: 1–4; Ulbrich and Werth 2021: 23–26). It is also clear that intra-speaker variation is distinct from inter-speaker variation and demands its own research methods and frameworks (Ulbrich and Werth 2021: 9–10).

It has been demonstrated that morphosyntactic intra-speaker variation is widespread and empirically robust (Bülow et al 2021: 1–4; Ulbrich and Werth 2021: 9–10). For example, Ellsäßer and Schallert (2021: 207–214, 226–230) demonstrated that speakers alternate between morphosyntactic variants (e.g., verbal complex word orders), but these alternations obey co-occurrence restrictions: certain variants cluster together within individuals. Based on these findings, the authors argue that intra-speaker variation should be modeled as probabilistic

constraint interaction, rather than optional rule application. Additionally, Nickel (2021: 178–185, 190–198) found intra-speaker variation in nominal inflection, even in elicited contexts. This study found that variation occurs within a single speaker’s speech and within the same grammatical context, contradicting the idea that inflection is categorically fixed per speaker. This evidence suggests that intra-speaker variation is structured and often predictable in distribution, even though it may look like ‘free variation’ on the surface.

Traditionally, it has been assumed that intra-speaker variation follows patterns that are also observed in inter-speaker variation. However, more recent research suggests that group-level syntactic variation cannot be straightforwardly projected onto intra-speaker variation (Ellsäßer and Schallert 2021: 214). The problem of the ergodicity⁴ assumption is discussed in several studies (Ellsäßer and Schallert 2021: 214; Bülow et al. 2021: 9–10, 35–36; Molenaar 2008), which point out that population averages are not reliable predictors of individual variation. This stance suggests that averages can mask individual differences and that morphosyntactic variation must therefore be studied at the individual level rather than inferred from community-level distributions.

⁴ Ergodicity is a mathematical principle that argues that the average outcome of a group is the same as the average outcome of the individual over time.

2 PAST PARTICIPLE LEVELING IN THE SBSCAE CORPUS

This chapter concerns the empirical part of the thesis. It describes both corpora and methodology, as well as specific examples from the corpora. Finally, it provides an overview of the findings. The first subchapter describes the two corpora used in the thesis: SBSCAE and COCA. The second subchapter describes the procedures of data extraction from the corpora and the methods used for the analysis. The third subchapter presents an analysis of all PP leveling cases found in SBSCAE, including a discussion of *have got* and *have gotten*. And, finally, the fourth subchapter presents the findings in a summarizing manner.

2.1 Corpora: SBSCAE and COCA

The research question of this thesis was: What are the meaning, usage contexts, and formal properties of the PP leveling construction in spoken American English? This thesis will attempt to answer the research question by examining the Santa Barbara Corpus of Spoken American English (SBSCAE) (Du Bois et al 2000–2005), identifying instances of PP leveling and normative present perfect, and creating schematic representations of the constructions within the CxG framework. Moreover, in one case, as there were no suitable examples for comparing the PP leveling construction with the standard one, two similar examples from the spoken part of the COCA corpus were provided.

SBSCAE consists of recordings of naturally occurring interactions from across the United States (Du Bois et al 2000–2005). It features acts of speech from a diverse range of individuals representing various regions, ages, professions, genders, and ethnic backgrounds. While most recordings in the corpus are face-to-face conversations, it also includes other types of spoken interactions, such as telephone conversations, card games, food preparation, on-the-job talk, classroom lectures, sermons, storytelling, town hall meetings, and tour-guide talks.

The corpus is also part of the International Corpus of English (ICE). The SBCSAE is approximately 249,000 words⁵ and features 213 speakers in 60 recordings.

The SBCSAE corpus was selected for this study for several reasons. Firstly, the corpus size seems ideal: neither too large nor too small, allowing for qualitative analysis with manual extraction while likely providing sufficient data to identify a rare phenomenon such as PP leveling. Secondly, the corpus metadata provides quite extensive descriptions of recording situations and of speakers and their English dialects/varieties (demographic information for the speakers of the SBCSAE corpus is discussed in section 2.3.1, Table 3). Moreover, the SBCSAE includes descriptions of event types for each recording (see Appendix 1), which provide crucial information about the contexts/styles/registers of conversations. 33 of 60 recordings in the corpus are face-to-face conversations, making this the prevalent event type in the SBCSAE. Other less frequent event types include task-related talks, sermons, teaching, discussions, etc. The speakers and topics in conversations in this corpus are also quite diverse, allowing the analysis of PP leveling across various discourse contexts.

It is also necessary to acknowledge some limitations of this corpus. The format of the corpus files is CHAT, a format developed for the CHILDES corpus of children's spoken language (MacWhinney 2000: 4–6) and now used across various projects as a transcription format. An example CHAT file is provided in Appendix 2. CHAT encoding format requires its own search system, CLAN. CLAN somewhat resembles Regular Expressions; however, it still uses its own syntax, making it quite challenging to learn and use. Moreover, CHAT files in the SBCSAE corpus also include annotations for phonological features and timing information, which may be redundant in this study and pose difficulties during extraction.

Another potential issue with the SBCSAE corpus is the timeframe of its compilation. The corpus was compiled between 2000 and 2005, and most of the recordings were made even

⁵ Numbers reported from the official description of the corpus, token count unknown.

earlier, in the 1990s. This time placement makes the corpus over 20 years old. Compared to COCA, a dynamic corpus with the most recent addition in 2019, the last part⁶ of SBCSAE is 14 years older. It is essential to recognize that language is constantly evolving, and numerous changes may have occurred during this period. However, it must also be noted that spoken corpora compilation is quite a labor-intensive and elaborate process that requires large amounts of time and resources, which is possibly one of the reasons why such a low number of spoken corpora exist today. Since readily available spoken English-language corpora that provide access to raw data are difficult to find, the SBCSAE corpus appears to be the most suitable option for this thesis.

The spoken part of COCA (Corpus of Contemporary American English) (Davies 2008) was used as a supplementary corpus in one case, where there were no similar standard constructions for comparison in the SBCSAE corpus. The spoken part of the COCA corpus is many times larger in size than the SBCSAE corpus – 127,396,916 words in total. The corpus comprises speech recordings made between 1990 and 2019. It mainly consists of broadcast recordings from TV and radio. The size of the corpus is the reason why it may be useful to draw comparative examples from, as it allows one to find many instances with rarer verbs. It is important to note, however, that the speech in the COCA corpus may be less spontaneous and more careful than the one in the SBCSAE corpus. Nevertheless, it should not affect the analysis in this thesis, as the corpus is used to identify the standard present perfect construction present in both more casual and more careful speech, as well as in spontaneous and scripted or elicited speech. The COCA corpus was used only once, whereas other example pairs were found from the SBCSAE corpus.

⁶ The corpus was published in four parts between 2000 and 2005.

2.2 Method of Analysis and Data Extraction

Extracting syntactic patterns can be methodologically challenging or overly complex, especially when the corpus lacks morphosyntactic annotation, as is the case with SBCSAE. Previous studies (e.g., Chatten et al. (2024) or Geeraert and Newman (2011)) solved this problem by first deciding on the list of irregular verbs that can occur in the present perfect construction, and then compiling a corpus query based on the potential verb forms. As discussed above (section 1.3.1), such a data-extraction method can miss certain hits (e.g., if the author did not predict a particular verb). It is generally accepted that accuracy may be compromised in quantitative corpus studies. Since the corpus of this study is significantly smaller, it was decided to combine automatic and manual extraction techniques to maximize accuracy. Additionally, the decision was made to avoid AI whenever possible out of a reluctance to sacrifice precision, even though it was eventually used for the analysis of the 's dataset due to the lack of other solutions. As a first step, all the concordance lines containing any form of *have* were automatically extracted. The dataset, containing approximately 13,394 concordance lines, was manually annotated to identify occurrences of the present perfect (both standard and leveled).

Since one of the aims of this thesis is to compare the standard present perfect with leveled constructions, it is necessary to establish what is considered standard. Following Geeraert and Newman's approach, this thesis will use the list of irregular verbs provided in Quirk et al. (1985: 115–120). The list of relevant verbs and their standard forms is available in Appendix 3. The standard meaning of the present perfect is also established following Quirk et al. (1985: 192–193). The author defines the present perfect as “relating a past event/state to a present time orientation” (Quirk et al. 1985: 192) and classifies it into three main meanings: (a) state leading up to the present (e.g., *That house has been empty for ages.*); (b) indefinite events leading up to the present (e.g., *Have you (ever) been to Florence?*); (c) habit (i.e., recurring event) leading up to the present (e.g., *Mr Terry has sung in this choir ever since he was a boy.*).

As a first step, the CLAN command interface was used to extract all the occurrences of *have* from the corpus. The list of CLAN commands is available in Appendix 4. These commands were used to find all instances of *'ve*, *have*, *has*, *haven't*, and *hasn't*. The instances of *'ve*, *have*, *has*, *haven't*, and *hasn't* were then analyzed by the author manually (total 13,394 lines) to determine whether they belong to a present perfect or a PP leveling construction based on the list of standard irregular paradigms identified in Appendix 3. The hits with irrelevant constructions, such as “have to do something” (e.g., *we're gonna **have to find somewhere to get something*** (example from the SBCSAE corpus)), were then manually removed from the dataset. Because the extracted data had to be reviewed manually, the procedure proved especially time-consuming.

In case of *'s*, which could be a shortened version of *is* or *has* and a possessive marker, the categorization was performed with the help of LLMs (or AI, used here interchangeably). A CLAN command that searched for *'s* was created (available in Appendix 4) to identify all instances of *'s* in the corpus. Since the number of hits exceeded 6500, manually classifying the occurrences proved challenging, so AI was employed for assistance. The search results were inserted into ChatGPT (using the GPT 5.3 Instant model), which determined whether the extracted instances of *'s* included any present perfect constructions. It is necessary to acknowledge that LLMs do not achieve 100% accuracy in analyzing language, any more than humans do. To ensure acceptable accuracy, a 10% sample (650 hits) was manually reviewed, and no errors were identified. Moreover, the analysis was double-checked and confirmed by another AI chat, Copilot (using GPT 5.5). Nevertheless, these precautions do not guarantee a human-like accuracy, but at least ensure that most of the output is acceptable. As a result, only one present perfect construction was found, which did not include PP leveling and therefore did not affect the analysis.

One possible shortcoming of the search method adopted in this study is that any non-standard versions of the present perfect, such as *ain't* instead of *have/has* or omission of auxiliary, may have been missed. This is not seen as a problem, as some of these forms may be considered different constructions and are therefore not considered directly relevant to this study. Moreover, to narrow the focus of the analysis to present perfect and PP leveling only, such constructions as passives (example (1)), present perfect continuous (example (2)), tag questions or elliptical constructions (examples (3) and (4)), non-standard perfect constructions with regular verbs (example (5)) were deliberately excluded from the dataset.

- (1) *We had to go some place for Hector's work first you know The Wonder Years. where I got insulted by this other little assistant who since **has been fired**.* (SBCSAE, SBC006.cha)
- (2) ***Has she been clicking** with your friends?* (SBCSAE, SBC028.cha)
- (3) ***Has it?*** (SBCSAE, SBC013.cha)
- (4) *Yes **she has**.* (SBCSAE, SBC011.cha)
- (5) *Now w- how- why **has** God **use** him.* (SBCSAE, SBC030.cha)

Files in the SBCSAE corpus are prosodically annotated. These include inhales, exhales, intonation, laughs, coughs, and pauses (example (6)). Since this study examines syntax rather than phonetics, these annotations were deemed irrelevant. The examples were edited to exclude such annotations. Time tags, as seen in example (7), were also manually removed from the examples because they were irrelevant to the analysis. Speaker annotations (example (8)) were removed from the analysis, as the speakers' pseudonyms are provided separately before each example and their demographic information is available in 2.3.1 (Table 3). Other stylizations, such as capitalized letters and punctuation, were not changed, but the spaces before punctuation were removed for ease of reading.

- (6) *[Green Bay] won twenty-four to nine [2teen]2 . [2 So now]2 we can]3 have a merry]3 Christmas [% laugh] .* (SBCSAE, SBC059.cha)
- (7) *(..) Well I know •149942_150679• but •150679_150888• [normally] you would give em three-day notice to perform uh . •150888_153139•* (SBCSAE, SBC053.cha)
- (8) **DOLO: You want paprika or chili . •28754_30381•
JULI: (..) Chili . •30381_31014• (SBCSAE, SBC037.cha)

The constructional analysis was presented as a schematic representation of the competing constructions. Several constructional representations were considered for this study (e.g., allostructions in Capelle (2006: 18), constructions with internal/external meanings in Pöldvere and Paradis (2019: 21), or Croft and Cruse's (2004: 247) more traditional representation). However, since this thesis is concerned with constitutional competition, the constructional representation was expected to reflect that. Thus, an adopted version of Croft and Cruse's construction structure in Figure 2 was chosen. Representations in Figures 5 and 7 are modeled based on Croft and Cruse's (2004: 252) representation of the *spill the beans* construction shown in Figure 4, where lowercase is used to show the form and uppercase is used to show the meaning. The PP leveling and present perfect constructions in Figures 5 and 7 were positioned side-by-side to represent competition.

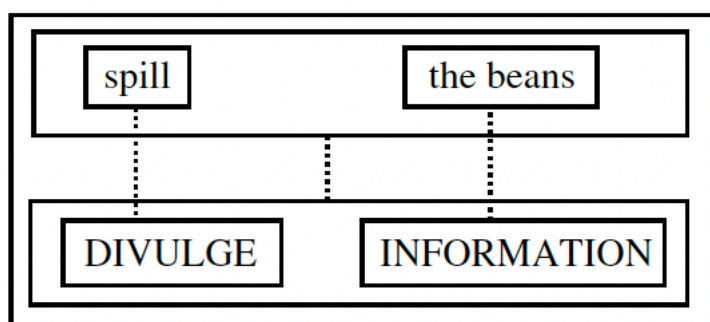


Figure 4. Constructional representation of *spill the beans* from Croft and Cruse (2004: 252).

Another factor considered in the analysis is the style or register of the speech act. The style was determined from the SBCSAE metadata (see Appendix 1) and modeled on a scale from most casual to most careful (see the scale in section 2.4, Figure 6). As discussed in section 1.4.2, style or register can be modeled as a continuous factor rather than a binary categorization. Given the nature of the SBCSAE, where event types can be understood as referring to the nature and style of the conversation, a continuous model of style seemed more suitable. The scale labels event types found in metadata, from most casual (i.e., more informal) to most careful (i.e., more formal).

2.3 Past Participle Leveling in SBCSAE

This section presents the results obtained from the SBCSAE corpus. There are eight sections in this subchapter. The first section provides an overview of PP leveling occurrences in the corpus and briefly compares them with previous studies. The following sections are titled according to the analyzed verbs and include qualitative analyses of all PP leveling examples found in the SBCSAE. The last section discusses the *have got* and *have gotten* occurrences in the SBCSAE.

2.3.1 Frequency and Distribution of Past Participle Leveling in SBCSAE

Table 2 presents the distribution of standard present perfect and PP-leveled occurrences across six datasets, expressed in absolute numbers. The total number of tokens for the PP-leveled construction is 9, the number for the standard present perfect construction is 809. Given the corpus size of 249,000 words, the normalized frequency of leveled constructions is 36.1 leveled constructions per million words. Comparing the number of occurrences of standard present perfect and PP leveling present perfect, it can be said that PP leveling constitutes only around 1% of all the cases of present perfect, and the standard construction makes up 99% of occurrences. Thus, PP leveling can be characterized as extremely rare and marginal in the present perfect in the SBCSAE corpus.

Table 2. Number of occurrences of standard present perfect and PP leveling in the SBCSAE corpus in absolute numbers in the six datasets.

Dataset	Standard present perfect	PP leveling	Total
<i>has</i> dataset	76	2	78
<i>hasn't</i> dataset	16	0	16
<i>have</i> dataset	126	3	129
<i>haven't</i> dataset	54	1	55
's dataset	164	0	164
've dataset	373	3	376
Total	809	9	818

Table 3 presents each occurrence of PP leveling by example number and file name in the SBCSAE, with pseudonyms for the speakers and their demographic information provided in metadata by Du Bois (2000–2005). Nine tokens were produced by seven speakers, with one speaker producing two tokens. In summary, the speakers are between the ages of 15 and 51. While most speakers are between 29 and 40 years old, there are two visible outliers: an adolescent (15) and a middle-aged adult (51). Among the speakers, there are four men and three women. Even though the sample size does not allow firm conclusions, the data suggest that both men and women use the leveled construction. Most of the speakers have higher education; however, this may reflect the overall bias toward more highly educated individuals in the corpus, as it was collected through the channels of the University of California, Santa Barbara. The speakers are also predominantly white, possibly also reflecting overall trends in the corpus. Speakers come from a range of professions and roles. The home regions of the speakers also vary across the corpus as a whole. The speakers who use leveled forms seem to mostly come from the Midwest. Nevertheless, given outliers such as Texas and Washington, no conclusions can be drawn about PP leveling being associated with specific regions.

Table 3. Demographic information for speakers who produced leveled present perfect constructions in SBSCAE.

file name + number of example	name	gender	age	hometown	home state	current state	education	years of education	occupation	ethnicity
SBC044 : (9), (21)	Lajuan	M	35	Indianapolis	IN	WI	B.A./B.S.	16	Retail Management	Black/African American
SBC021 : (13)	Walt	M	40	-	-	MA	A.A	16	Evangelist and minister of the gospel	White
SBC039 : (14)	Kirsten	F	29	Holland (MI)	MI	IL	B.S.	16	Education	White
SBC015 : (16)	Ken	M	51	midwest	IN	IN	PhD M	23	Director	White
SBC030 : (18)	Bradley	M	39	S.E. Texas	TX	IL	M.A.	-	Baptist Minister	White
SBC040 : (19)	Debra	F	30	Louisville	KY	KY	B.S.	16	Horse Park Employee	White
SBC042 : (24)	Kendra	F	15	Tri-Cities	WA	WA	Some High School	10	Student	White
SBC026 : (12) ⁷	-	-	-	-	-	-	-	-	-	-

Table 4 summarizes verb lemmas and their classes, as well as co-occurrence with modals and first-person subjects. The vast majority of verbs in the leveled present perfect belong to class (e) (paradigm shift from class (e) to class (b)), with one exception, *beat*, which belongs to class (c) (paradigm shift from class (c) to class (a)). Surprisingly, modals are not as frequent, appearing in only three of the nine examples. The first-person subject appears more often, with six of nine examples. It also seems that when a modal is present, then the first-person is also present in two out of three cases. What is also striking is that only the class (c) verb (*beat*) lacks both modal and the first-person subject, although it may be coincidental.

⁷ No speaker information is available due to the nature of the recording – public speech.

Table 4. Verb lemmas, verb classes, and co-occurrence with modals and first-person subject for each leveled example.

example number & verb phrase	verb lemma	verb class	modal (yes/no)	first-person subject (yes/no)
(9) <i>have chose</i>	choose	(e)	yes	yes
(12) <i>have chose</i>	choose	(e)	no	yes
(13) <i>have chose</i>	choose	(e)	no	yes
(14) <i>have went</i>	go	(e)	yes	no
(16) <i>have went</i>	go	(e)	no	yes
(18) <i>have saw</i>	see	(e)	no	no
(19) <i>have beat</i>	beat	(c)	no	no
(21) <i>have spoke</i>	speak	(e)	no	yes
(24) <i>have took</i>	take	(e)	yes	yes

Table 5 presents the event types (i.e., conversation types/contexts) for each example in the SBCSAE. In summary, the most frequent conversation type is ‘conversation between friends’, which appears three times. This type of conversation is the most frequent in the corpus overall. A closely following second is ‘sermon’, which appears twice.

Table 5. Event types by example number.

example number & verb phrase	event type
(9) <i>have chose</i>	conversation between friends
(12) <i>have chose</i>	city meeting
(13) <i>have chose</i>	sermon
(14) <i>have went</i>	city meeting
(16) <i>have went</i>	conversation between friends
(18) <i>have saw</i>	sermon
(19) <i>have beat</i>	scripted tour
(21) <i>have spoke</i>	conversation between friends
(24) <i>have took</i>	family argument

Having discussed the general frequency and distribution of PP leveling in SBCSAE, the following sections will focus on the qualitative analysis of each example separately. The aim of this analysis is to discuss the form and meaning of the leveled construction, as well as any influences of the context in which the construction occurs.

2.3.2 *Have chose*

Example (9) demonstrates an excerpt from a conversation between two friends in Milwaukee, Wisconsin, where one friend, with a pseudonym Lajuan, talks about their experience of being gay. Lajuan is 30 years old and originally from Indiana. The comparative standard examples (10) and (11) were taken from the COCA corpus and are similar in context, as the conversations also involve topics of sexuality.

(9) *uh what it is is that because of what society has done or has laid the rules for us I don't really think I would **have chose** to be gay.* (SBCSAE, SBC044.cha)

(10) *I know that Mel did not have a choice about his sexuality, and if he did have a choice, he would **have chosen** to be married and have a family and be with me.* (COCA, spoken)

- (11) *It's as simple as making a choice. You've **chosen** to be gay. You can choose not to be.* (COCA, spoken)

Firstly, it is necessary to identify what constructions overlap in this utterance. This utterance uses the modal perfect construction, which encompasses a perfect construction. The construction 'have chose' has been reported to co-occur with modals 39% of the time by Geeraert and Newman (2011: 16). A standard form of the modal perfect construction would be [would have + Vpp] as seen in example (10), but here in example (9), it is [would have + Vpt]. Regarding the form of both constructions, it is worth noting that the modal construction may influence how the verb is used, as discussed in section 1.3.3. The construction of a modal with non-finite perfect does not have the same syntactic structure and does not retain exactly the same meaning as the finite perfect construction. It can be argued that the non-finite perfect is part of the larger modal construction. Examining example (9), it is evident that the meaning of *would have chose* is hypothetical past. While the construction found in example (11) clearly signals the resultative or continuous nature of the action, there is no such meaning in the modal construction in example (9). Moreover, as discussed in section 1.3.3, Chatten et al. (2024) suggested a reanalysis of the [modal + non-finite perfect] construction, which depends on the phonetic reduction of *would have* into *woulda*. Even though the authors of the SBCSAE corpus transcribed the auxiliary as *have* and not 've, after listening to the original recording, it can clearly be heard that *have* is quite reduced, as the speed of the utterance is relatively high, making the *woulda* reanalysis possible.

As Chatten et al. (2024: 28) found a correlation between the subject person and PP leveling in one of their US corpora, let us examine the person of the subjects in examples (9) through (11). In example (9), where leveling occurs, the subject is the first-person pronoun. In examples (10) and (11), where no leveling occurs, the subjects are third-person and second-person, respectively. Such a pattern matches that reported in Chatten et al. (2024), and it is likely that the subject person triggers the leveling in example (9).

Nevertheless, the other two cases of PP leveling with the verb *choose* do not include modals. Example (12) presents a sentence that contains PP leveling, taken from a transcription of a city meeting in Chicago, in which officials spoke with members of the public about a grant to fund community development. The speaker information is not available for this recording, as it was a public meeting.

(12) *I **have chose** not to do it but out of necessity.* (SBCSAE, SBC026.cha)

This construction follows the pattern [have + Vpt]. In this case, there are no overlapping constructions influencing the verb form choice. Examining the conversational context of this utterance and the whole interaction, it is clear that the style is more careful than in example (9), but still conversational. Compared to example (9), where the conversation was between friends, this is an official meeting, where the language may be more formal. Nevertheless, the non-standard construction is present here. The register alone does not appear to determine the presence of PP leveling. Similarly to the previous example, example (12) also includes a first-person subject, which seems to exemplify the previously found correlation between the first-person and PP leveling.

Another example of *have chose* comes from a sermon, recorded in Boston, Massachusetts, that was described in the metadata of SBCSAE as ‘rather lively’. The speaker is Walt, aged 40, whose English dialect and state of origin are unknown.

(13) *But what **I've chose** is to overcome my fears and give my life to Jesus Christ.* (SBCSAE, SBC021.cha)

In example (13), the construction also takes the form [have + Vpt]. There is no modal construction that influences the leveled construction. As no modal influences the choice of verb form, it may be that leveling was prompted by the subject person in this example. The subject is in first person, so in accordance with previous findings, it may influence the leveling.

Interestingly, the discourse context is quite distinct from the previous two examples. A sermon is a specific type of speech that can be considered a register in its own right.

It is also worth noting that the verb ‘choose’ is a class (e) verb (see Figure 1) with three distinct forms for infinitive, past tense, and past participle in the paradigm. If we consider example (13) and determine the paradigm, it would be *choose-chose-chose* rather than the normative *choose-chose-chosen*. This new paradigm would fall into class (b), which is in line with the theory that class (b) attracts other verbs. Such behavior may exemplify the paradigm economy principle (see Carstairs 1983), where a language always strives for the logical minimum of forms within its paradigms.

2.3.3 *Have went*

The next most common PP leveling construction in the corpus is *have went*, with two occurrences. Out of two occurrences, *have went* appears in a modal construction one time. Let us examine example (14), an extract from a training meeting at an aquarium in Chicago, featuring an utterance by the 29-year-old speaker Kirsten, who is originally from Michigan. Example (15) is presented for comparison and comes from a conversation between friends, which may be more casual than the training meeting. The utterance comes from the speaker Miles, for whom there is little information other than that he⁸ is African American and currently in California.

(14) *The reason I mentioned that was if one was smaller he would **have went** for the smaller egg.*
(SBCSAE, SBC039.cha)

(15) *In fact I would **have gone** today if it wasn't for my back.* (SBCSAE, SBC002.cha)

In both examples, the PP leveling construction [have + Vpt] occurs within a modal construction [would have + Vpt], forming the construct [would + have + Vpt]. Additionally to the previously proposed explanations, it can be suggested that the presence of a modal creates an ‘auxiliary

⁸ Gendered pronouns are used for speakers when the metadata provides gender information.

stack' that increases the structural distance between *have* and the lexical verb (Vpt). From a usage-based perspective, such environments may increase processing demands and reduce speakers' sensitivity to the distinction between past tense and past participle. The current data are too limited to test this tendency accurately, but since the pattern has been reported in previous studies, this example may exemplify it.

As discussed in sections 1.3.3 and 2.3.2, Chatten et al. (2024) argued for the possibility of reanalyzing the modal perfect construction as [adverb + finite verb] (e.g., *I woulda went*). However, the condition for such a reanalysis is a drastic reduction of *have*, where it is reanalyzed as a new unit with the modal verb, i.e., *woulda*. Here, the transcription suggests less reduction, as the auxiliary is written out as *have*, not 've. However, after listening to the recording, there seems to be some reduction in the auxiliary, but certainly not to the same degree as in example (9). In the author's judgment, the reanalysis is still not possible here, as the /v/ sound can clearly be heard between the modal and the main verb, signaling the presence of the auxiliary *have*. It seems plausible that, in this particular case, the reason for leveling is not the reanalysis of the construction.

As person marking was found to influence PP leveling, let us examine the subjects in both examples (14) and (15). Interestingly, the subject in example (14), which exhibits PP leveling, is in the third person. Even though Chatten et al. (2024: 28) found that third-person subjects inhibit leveling in some cases in American English (see section 1.3.1), in examples (14) and (15), we see the opposite. Since Chatten et al. (2024) did not confirm that this tendency prevails throughout all American English, examples of the opposite are natural.

Example (16) comes from a conversation between three friends recorded in Los Angeles, California, so the register is quite informal. The utterance was produced by a speaker with the pseudonym Ken, who is 34 years old and from California. Example (17) is presented for comparison, and is from the same file and uttered by the same speaker as example (16).

(16) *I've **went** to the doctor.* (SBCSAE, SBC015.cha)

(17)[...] *there's delegations going every week thousands **have gone** [...].* (SBCSAE, SBC015.cha)

The central observation in examples (16) and (17) is that they exemplify intra-speaker variation, where a speaker uses both the prescriptively standard present perfect and the PP leveling construction. Because the two utterances (16) and (17) come from the same speaker, they cannot be attributed to regional variation, dialectal differences, or educational level. Instead, the examples may suggest that the two constructions exist in the speaker's repertoire simultaneously. From the Construction Grammar perspective, such alternation is consistent with constructional competition. Both [have + Vpp] and [have + Vpt] initiate the same higher-level construction of present perfect, but they differ lexically. The fact that Ken produces both variants within the same conversation indicates that the competition between *went* and *gone* for the participle slot is active.

However, since this is only one example and one comparison of Ken's speech, there is no way to know whether such alternation is systematic for him. Other explanations are also possible. For example, a more generative analysis may argue that this is merely a slip of the tongue, meaning that the speaker's internal grammar licenses only the construction [have + Vpp], but performance errors during production produce such constructions as [have + Vpt]. Since earlier studies also report intra-speaker variation, the conclusion drawn in the present study leans towards intra-speaker competition between constructions rather than performance errors.

Examining the subject person, the observed dependency of PP-leveled constructions on first-person subjects is consistent with earlier studies. In example (16) that exhibits PP leveling, the subject is first person. However, in the standard usage (example (17)), the subject is in the third person. It is, therefore, possible that the choice of an alternation is conditioned by the subject person for this speaker. Although again, since we only have access to one instance of variation in this speaker's speech, it is possible that the choice is conditioned by other factors.

2.3.4 *Have saw*

The PP leveled construction with the verb *saw* appeared only once in the SBCSAE corpus, in example (18). This example appears in a recording of a sermon delivered at a Baptist church in Chicago, Illinois. The speaker's pseudonym is Bradley, and he is 30 years old. Bradley is from Texas and has a Texan dialect.

(18) *But God **has saw**- seen fit to use us in that way.* (SBCSAE, SBC030.cha)

Example (18) is particularly revealing, as it demonstrates real-time self-repair. The speaker initially produces the PP leveling present perfect (*has saw*) but immediately corrects it to the standard realization (*has seen*). This may exemplify real-time competition between the alternative realizations of the present perfect construction.

The context of a sermon is relatively formal, which may encourage the speaker to adhere to the standard construction, thereby triggering self-correction. Rather than showing a stable grammatical system, this example may suggest that the variation is entrenched in the speaker's mind. The self-repair may suggest that under conditions of higher monitoring (in a more careful style or formal context), the speaker is choosing between the two competing constructions. Since the corrected construction is standard in this instance, it can indicate that standard constructions are preferred for a more careful style. Nevertheless, it is also possible that this is an instance of a performance error. Although such an explanation is less likely according to usage-based theories, an error during production may have led to a leveled form, such as *has saw*, appearing in the speaker's speech.

What is also striking about example (18) is its relation to previously reported influences on PP leveling. The subject of the sentence is in the third person, which was reported to inhibit PP leveling. Moreover, such verb paradigms as *see-saw-seen* are also rarer in PP leveling, as the phonological difference between the past tense form and the past participle is greater. One

possible explanation could be that self-correction occurs because unfavorable conditions affect the speaker's judgment of the acceptability of PP leveling in this context.

Focusing on the lexical verb *see* in example (18), we observe that the verb undergoes a shift from the paradigm *see–saw–seen*, to *see–saw–saw*. Originally a class (e) verb, *see* behaves like a class (b) verb (see section 1.2, Figure 1). This exemplifies the broader tendency of class (e) verbs to adopt the class (b) paradigm.

2.3.5 *Have beat*

Have beat appears in the corpus once in example (19), which was extracted from a transcription of a scripted tour recording from the Kentucky Horse Park. The speaker is 30-year-old Debra, whose dialect region is established as Kentucky. The speaker's intonation in the audio recording suggests she is reading from a source rather than speaking spontaneously or reciting from memory, which may affect her language use.

(19) *Since then Alisheba **has beat** his record but John is still the richest gelding.* (SBCSAE, SBC040.cha)

(20) *The jockey that rode Forego the great Willy Shoemaker said there wasn't a horse alive that **could've beaten** Forego on that day.* (SBCSAE, SBC040.cha)

Comparing examples (19) and (20) from the same recording produced by the same speaker, an intra-speaker variation can be observed (see also examples (16) and (17)). Nevertheless, PP leveling occurs only once in this speaker's speech throughout the recording. The present perfect construction is used only twice: once in example (19) with PP leveling and once in example (20). Previously, it was mentioned that the PP leveling phenomenon tends to co-occur with modals (see examples (9), (10), (16), and (17)). However, examples (19) and (20) demonstrate an opposite dependency, where the same speaker chooses to use the past tense without a modal, and the past participle with a modal. This may reflect the differences between lexical verbs in their tendency to co-occur with modal verbs. Referring back to Geeraert and Newman (2011: 16), the leveled verb *beat* has a 31% rate of co-occurrence with modals. Compared to the 39% rate of the verb *choose* (in examples (9) and (10)), it is 8% lower. Another

key observation is that the auxiliary in example (20) is quite reduced, as, upon listening to the recording, there appears to be no or very slight /v/ sound after the modal, with the whole utterance sounding similar to '*coulda beaten*'. Even though this is the case, the construction is not leveled, so the reanalysis probably did not occur here.

From the audio recording, it can be heard that the speaker is reading rather than speaking spontaneously or reciting from memory. Therefore, some inconsistencies with earlier studies can be explained by the difference in discourse type. Nevertheless, it is surprising that both *beat* and *beaten* as past participles appear in a single instance of scripted speech. This said, *beat* is somewhat different from other verbs discussed in this thesis, since *beat* as a past participle variant has been listed already in Quirk et al. (1985). Thus, *beat* may be perceived as more standard than instances of leveling in other past participles.

What makes example (19) more distinct than the other examples is the shift in verb paradigms observed here. The standard paradigm of the verb *beat* is *beat-beat-beaten*, which makes it a rare (c) class verb – only 0.6% of irregular verbs in English fall into this category (see section 1.2, Figure 1). In example (19), *beat* is used as a past participle, shifting the paradigm into *beat-beat-beat*. This paradigm should be allocated to class (a). Unlike all other examples in this analysis, the shift happens from class (c) to class (a). The shift from class (c) to class (a) may happen more easily, as the phonological difference is less significant between the past simple and past participle in this class. In the case of *beat*, the past participle differs from the past simple form only by a suffix. Compared to, for example, *write-wrote-written* that both adds the suffix and changes the vowel in its past participle, *beat-beat-beaten* shows less of a difference.

Moreover, the subject's person also needs to be discussed, as it may influence the construction. In example (19), the subject is in third person, which was found to inhibit leveling.

Nevertheless, the leveling co-occurs with the third person, so it does not seem to influence the leveling here.

Considering the meaning of the PP-leveling construction, it can be said to be consistent with the standard present perfect meaning. The past event in example (19) continues to influence the present; thus, the aspect can be characterized as resultative, and both constructions seem to share the same meaning here.

2.3.6 *Have spoke*

Have spoke also appears in the corpus once, in the same recording as example (9), produced by the same speaker, Lajuan, who is originally from Indiana. The setting is a living room conversation between two friends in Milwaukee, Wisconsin. The transcription of the utterance is presented in example (21).

(21) *I haven't spoke to my brother to be honest with you since.* (SBCSAE, SBC044.cha)

What should be noted about example (21) is that the auxiliary *have* is negated. Negation was previously found to increase the PP leveling in one of the American corpora in Chatten et al. (2024: 28). While this does not mean that this is characteristic throughout American English, as Chatten et al.'s (2024) results can only be interpreted as indicative rather than conclusive, it is still necessary to acknowledge that negation has been observed to occur with PP leveling in example (21) and it may have induced the PP leveling in this case. Moreover, example (21) demonstrates co-occurrence of the leveled form with the first-person subject, which was also found to induce leveling.

Example (21) is particularly intriguing, as both utterances in example (9) and example (21) were produced by the same speaker. While in example (9), the PP leveling could be explained by the presence of a modal, considering example (21), it can be proposed that the discussion of individual mental grammar is of importance. The recurrence of PP leveling within the same recording by the same speaker makes this case less likely to be interpreted as a

performance error during production, but rather suggests a reflection a more systematic feature of the speaker’s mental grammar. In examples (16), (17), (19), and (20), we observed intra-speaker variation once. Here, the speaker may have the PP leveling construction more entrenched in his mind, as he uses it twice in the same recording that lasts only 29 minutes.

At the same time, it should be noted that Lajuan uses the standard present perfect construction in several other instances, for example, in examples (22) and (23).

(22) *I’ve **known** since I was gosh probably I can remember? Four?* (SBCSAE, SBC044.cha)

(23) *I’ve always **flown** everywhere by myself.* (SBCSAE, SBC044.cha)

Since this is the only speaker who uses PP leveling more than once, it might be useful to examine his verb usage across the whole conversation, as it may be argued based on the conservatism via entrenchment principle (Goldberg 2019) and Chatten et al.’ (2024) findings about frequency tendencies that the least frequent verb lemmas in his speech would be the ones used with PP leveling. The verb classes and lemmas used by Lajuan can be seen in Table 6.

Table 6. Verb classes and lemmas in Lajuan’s speech.

Class	Lemmas	Tokens	Forms and Counts
class (e)	<i>be, do, fly, know</i>	10	been ×7, done ×1, flown ×1, known ×1
class (b)	<i>have, deal, meet, tell</i>	6	had ×2, dealt ×2, met ×1, told ×1
class (d)	<i>become</i>	1	become ×1
regular	<i>travel, talk</i>	2	traveled ×1, talked ×1
get/gotten	<i>get</i>	2	gotten ×2
PP leveling (class (e) → class (b))	<i>speak, choose</i>	2	spoke ×1, chose ×1

At first glance, it is noticeable that class (e) and class (b) verbs are used the most; however, it must be noted that most of these verbs are extremely frequent verbs that often function as copulas, e.g., *be* and *have*. Due to the high token frequency of such verbs, the effect of conservatism via entrenchment may apply, i.e., the construction is less susceptible to innovation due to high form entrenchment in the speaker’s mind. Removing such verbs from

consideration, we can see that Lajuan uses class (e) verbs three times and class (b) verbs four times in standard constructions, which is nearly balanced. This means that there is no explicit evidence for the preference of one verb class over another.

Examining the verb in example (21), we observe the paradigm *speak-spoke-spoke*, which is classified as class (b), whereas the standard paradigm is class (e) (*speak-spoke-spoken*) (see section 1.2, Figure 1). Similarly to most examples in this study, the shift happens from class (e) to class (b). As the difference between the past tense and past participle is only in the addition of the suffix *-en*, the shift may happen more easily in this paradigm. Moreover, the meaning of the present perfect in example (21) seems consistent with the meaning of the standard form. Here, it shows the resultative nature of the action, as the past event continues to influence the present.

2.3.7 *Have took*

Have took appears in the corpus only once in a family argument about one of the daughters staying the night at a friend's house without permission, recorded in Pasco, Washington, presented in example (24). The utterance is produced by one of the teenage daughters with the pseudonym Kendra, who is 15 and from Washington.

(24) *I think if I spent the night I **would've took** a shower.* (SBCSAE, SBC042.cha)

(25) *Since when **has** she **taken** her punishment* (SBCSAE, SBC042.cha)

Firstly, it must be noted that example (24) includes a modal, similarly to example (9). Reviewing the recording, the auxiliary sounds quite reduced and is consistent with the reanalysis as *woulda*. Thus, this construction may have been reanalyzed as [adverb + finite main V], explaining the occurrence of leveling. The comparative example (25) is also consistent with this claim, as it lacks a modal and has a tensed auxiliary. Consequently, the instance of PP leveling in example (24) may be induced by the reanalysis of the modal and the auxiliary.

Additionally, examining the subject person in both examples reveals consistency with previously reported tendencies. In (24), the subject is first-person; in (25), it is third-person. This is in line with the findings of previous studies, which suggest that the first-person seems to induce leveling, while the third-person inhibits it. All in all, the linguistic context seems unfavorable for the PP leveling in (25), compared to (24).

Example (24) is an utterance produced by one of the daughters of the family mentioned in the description of the recording. One interpretation could be that, since the speaker is a child, this instance of PP leveling is simply an error due to children's command of language sometimes being incomplete. However, the metadata states that all the children involved in the recording are teenagers. Because the speaker is older, errors associated with limited language competence in youth are very unlikely. Since the utterance was produced by a teenager, such language use may be indicative of the linguistic trends among younger generations. Limited data does not allow for drawing conclusions based on this single utterance.

To discuss the meanings of the PP leveling and the standard present perfect construction, example (25) was added for comparison. The utterance in the example (25) comes from the same conversation but is attributed to another speaker. Examining the meaning of the leveled construction in example (24) requires a discussion of the modal. As already mentioned in sections 2.3.1 and 2.3.3, when the present perfect appears with a modal, it loses its aspectual meaning of a result, influence on the present, or experience. The same can be said about example (24). The meaning of the construction in example (24) is hypothetical past; the hypothetical past event has no influence on the present and does not reflect an experience. However, in example (25), the standard present perfect can be characterized as referring to durative meaning signaled by the word *since*, i.e., the action started in the past (accepting the punishment) and continues now. Such a difference in meaning does not contradict the earlier idea that modals may affect the form of the non-finite present perfect within the construction.

Considering the verb ‘take’, Geeraert and Newman (2011: 16) state that the leveled form of *take* co-occurs with modals 39% of the time. Interestingly, the percentage is the same as for the verb *choose*, discussed in example (9). Examples (9) and (24) seem to be very similar. The verb *take* is a class (e) verb (*take-took-taken*) that shifts to class (b) (*take-took-took*) in PP leveling.

2.3.8 The Special Case of *have got* and *have gotten*

It is unclear whether this construction can be considered a PP leveling construction, since there is evidence that ‘have got’ has become the standard construction, especially in British English. Nevertheless, it seems that the variation is also often connected to the meaning of the construction. In American English, *have got* is preferred for possession, while *have gotten* may be used for ‘obtain’. It is thus necessary to examine the constructions in the SBCSAE corpus to determine whether such a distribution is reflected in the data. Table 7 shows the number of occurrences of *have got* and *have gotten* sorted by their meaning, either ‘obtain’ or ‘possess’. The category named ‘other’ signifies other meanings that could not be described by ‘obtain’ or ‘possess’. Most often, this category included the meanings of necessity (e.g., *I’ve got to go*) or change of state (e.g., *She’s gotten smarter*).

Table 7. Number of occurrences of ‘have got’ and ‘have gotten’ in the SBCSAE corpus, categorized by meaning.

construction/meaning	‘possess’	‘obtain’	other	total
<i>have got</i>	116	3	37	156
<i>have gotten</i>	0	6	17	23

Examining Table 7, it can be said that ‘have got’ and ‘have gotten’ in the SBCSAE corpus are compatible with the previous classification of the two forms. While ‘have gotten’ is clearly used only with the dynamic meaning, there are 3 cases of ‘have got’ used with dynamic

meaning instead of the supposedly preferred stative one. There may be several reasons for this. According to Anderwald (2020: 197), the dynamic and stative meaning can often be ambiguous and difficult to distinguish, especially in the lack of context. A sentence such as “I have got a car” may mean either “I obtained a car” or “I possess a car”. Moreover, Anderwald (2020) argues that the most prominent reason for the emergence of ‘have gotten’ in American English is the criticism that ‘have got’ was receiving in American prescriptive grammar. Anderwald (2020) also mentioned that for a period of time, ‘have got’ was used with a dynamic meaning – thus, the three exceptions in SBCSAE may be remnants of that use. A simpler reason may be confusion over the forms. While in British English it is clear to speakers that ‘have got’ is preferred, in American English, there may be different understandings of the preferred contexts for each form.

Another interesting observation is that ‘have got’ is used much more often than ‘have gotten’ – 156 occurrences of ‘have got’ vs. 23 occurrences of ‘have gotten’. Even though the meanings are distributed accordingly, it may be preferred to express the meaning of ‘obtain’ with other forms. Even though many sources state that *gotten* was revived in American English, my data suggest that, although it is present, its frequency is fairly low. *Have got* constituted approximately 87% of all the appearances and *have gotten* only approximately 13%. This may indicate that the revival and spread of *gotten* is still underway, and the form may still be marginal.

One thing is clear – *have got* and *have gotten* may be considered a PP leveling that has already happened, at least in British English. As ‘gotten’ gradually became obsolete, the verb ‘get’ shifted paradigms from *get-got-gotten* (class (e)) to *get-got-got* (class (b)). In British English, the competition between ‘have got’ and ‘have gotten’ was resolved by language change, where one of the forms was dropped (see the construction grammar and variation chapter). However, the same competition, revived in American English by prescriptive criticism of ‘have

got', was resolved through specialization, with each form acquiring new usage contexts. This example confirms the possibility of such a shift occurring again, as may be the case with other verbs undergoing PP leveling. Since the 'have got' and 'have gotten' forms have already undergone the process of PP leveling and language change, they will not be discussed as PP leveling in this thesis.

2.4 Discussion

The research question of this investigation was: What are the meaning, usage contexts, and formal properties of the PP leveling construction in spoken American English? The findings suggest that PP leveling in spoken American English is a rare variant of the present perfect. Its meaning largely overlaps with that of the standard construction, except for the modal perfect construction, which will be discussed separately. The form of the leveled construction reflects a simplification of the paradigm, and its usage contexts span a wide stylistic range. The Discussion section answers the research question, summarizes the thesis findings in a synthesized manner, and relates them to previous research. This section is organized according to the three aspects of the research question: meaning, formal properties, and context.

Discussing the meaning and formal properties of the leveled construction, the observed alternation can be interpreted as normative constructions competing with leveled constructions in speakers' grammars, particularly in light of previous studies and intra-speaker variation. Figure 5 demonstrates both the semantic and formal properties of both constructions. It is proposed that the two constructions are in competition, as they have the same meaning but different forms. The examples in the thesis are consistent with an analysis in which the standard and leveled present perfect constructions function as competing options within the same constructional domain, rather than result from random production variability. As discussed in section 1.4.2, this competition may be resolved in the future through forms acquiring different

meanings or through new usage contexts. Let us, therefore, also discuss the usage contexts found in SBCSAE.

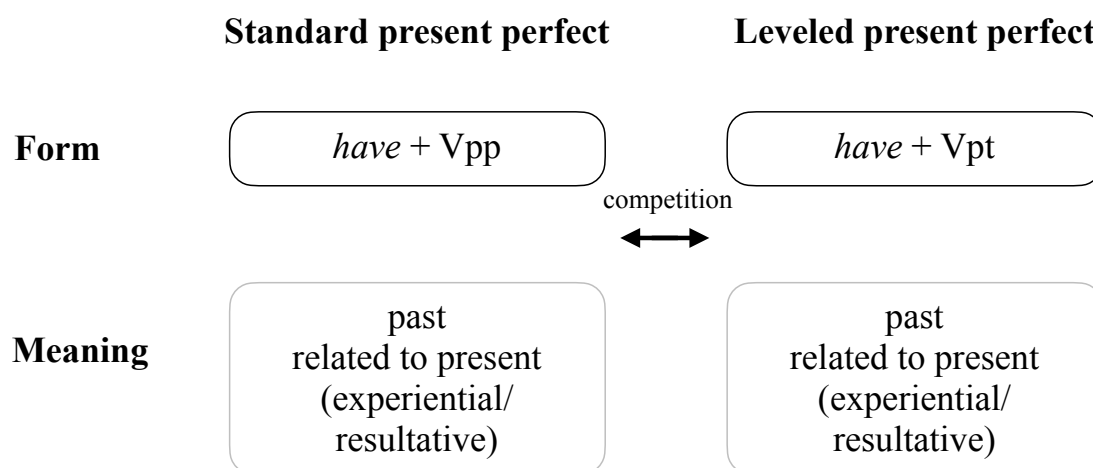


Figure 5. Schematic representation of the forms and meanings of the competing standard and PP leveling constructions.

Examining the contexts in more detail, we can observe that PP leveling has occurred across contexts with different levels of formality. In the SBCSAE, PP leveling was found in both casual conversations and more formal⁹ contexts. The register of contexts was modeled as a continuum, as shown in Figure 6 below. The figure models all the contexts of PP leveling found in the SBCSAE, ranging from more casual to more careful. The number of occurrences for each context is indicated below the axis. The number in red is the occurrence that included self-repair. All the event types on which the contexts were modeled can be found in Appendix 1. As shown in Figure 6, PP leveling was observed across a range of contexts, from casual conversations between friends to scripted speech¹⁰ and more formal meetings. However, it is worth noting that certain contexts involve more cases of PP leveling – such as conversations between friends and scripted speech. Due to the rarity of PP leveling and the small size of

⁹ Casual/informal and careful/formal are used interchangeably.

¹⁰ Sermons were considered scripted speech here.

SBCSAE, this may be a non-significant correlation. However, since the corpus consists predominantly of face-to-face conversations, yet the phenomenon was also observed in much less represented conversation styles (see Appendix 1 for event type distribution), it may be that there is no connection between the level of speech spontaneity and the preference for leveled forms. Such tendencies must be tested in a future quantitative study.

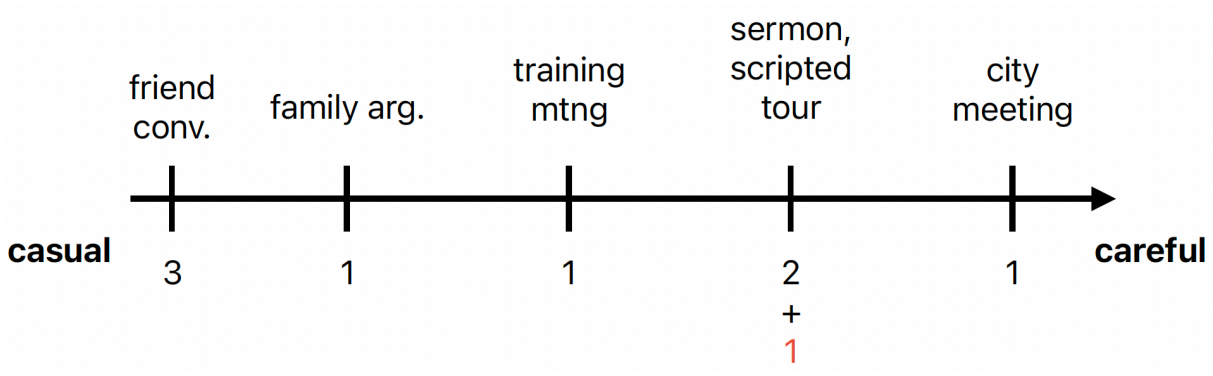


Figure 6. Contexts in which PP leveling occurred in the SBCSAE, modeled from most casual to most careful, with the number of occurrences below the axis.

Even though slight tendencies related to context have been noted in the SBCSAE, PP leveling slightly leaning towards more casual contexts, the leveled construction was observed at both ends of the scale. As no previous studies examined conversational contexts qualitatively, this finding is a new addition to the discussion of the phenomenon. This may also show constructions in competition in many distinct contexts at the time of corpus compilation. It must be noted that the corpus is quite old, and the patterns of stable variation and competition may have changed since the early 2000s. As such, a corpus study with a more recent dataset would be beneficial for the discussion. It is yet to be seen whether this competition will result in language change or specialization.

To further the discussion of possible competition and relation to previous findings, we must also describe verb paradigm shifts found in the corpus. Table 8 presents each case of PP leveling, including the verb lemma and the shift in paradigm. Table 8 shows that the

overwhelming majority of verb paradigm shifts in SBCSAE were from class (e) to class (b), as was predicted based on previous research in section 1.2. However, one case of the shift from class (c) to class (a) has also been found with the verb *beat*. Thus, the results here are compatible with the hypothesis that the most affected paradigm is class (e) and the most attractive verb class is (b).

Table 8. Verb lemmas and paradigm shifts found in SBCSAE.

paradigm shift	verb lemmas	number of occurrences
class (e) → class (b)	<i>choose, go, see, speak, take</i>	8
class (d) → class (b)	-	0
class (c) → class (a)	<i>beat</i>	1

Moreover, it is argued in this thesis that *have got* and *have gotten* are cases of PP leveling that have already happened not once, but twice, with different outcomes. Previous studies have determined that *have got* has become the normative variant of the present perfect construction with the verb *got* in British English. Historically, the British paradigm for the verb *get* was *get-got-gotten*. However, with time, the form *gotten* disappeared – the past tense won the competition of the two constructions and language change occurred. Due to certain prescriptive criticism of *have got* in the US, *gotten* was much later revived in American English, thus renewing the competition between the two constructions. Interestingly, this time, the competition was resolved in specialization: *have got* received the static meaning of ‘possess’ and *have gotten* the dynamic meaning of ‘obtain’. These cases of PP leveling are a striking example of the mechanisms behind competing constructions and language change. This thesis argues that the constructions investigated here are possibly in competition and may specialize or bring about language change in the future, similarly to what has already happened with *have*

got and *have gotten*. The case of *have got* and *have gotten* also demonstrates that it is important to examine the tendencies of individual verbs when studying PP leveling.

The frequency of the construction was found not to contradict previous findings (i.e., it was quite rare in the corpus) but to differ in normalized frequencies. It was found that PP leveling has a very low frequency: only 9 out of 818 present perfect tokens were PP leveling, that is 36.1 leveled constructions per million words, or around 1%. Geeraert and Newman (2011: 6) report that the number of occurrences per 10 million words in the spoken part of BNC was 184, that is, 18.4 PP-leveling constructions per million words. The number of occurrences of PP leveling in the SBCSAE corpus is almost twice higher than that in the spoken part of the BNC corpus. This comparison should, however, be taken with caution, as it is based on an extremely low number of tokens and is highly sensitive to corpus size.

Such a difference in frequency can potentially be explained by several things. Firstly, the difference in English variety. As Geeraert and Newman (2011) do not provide any statistics for the COCA, the only point of comparison is the BNC corpus, which represents the British variety. It is possible that PP leveling is simply more common in the American variety, even though Geeraert and Newman (2011) did not mention it in their study. The internal composition and balance of the corpora, especially regarding register, may also influence the drastic difference in frequencies. Secondly, given the small number of raw PP leveling occurrences in the SBCSAE corpus, the normalized frequency is extra sensitive to minor changes in token counts, which may exaggerate the difference in frequency between the two corpora. As COCA was compiled more recently than SBCSAE, the difference in frequency could also be connected to that. And finally, there are methodological differences between this thesis and Geeraert and Newman's (2011) study. This thesis attempts to capture all cases of PP leveling in the present perfect in SBCSAE by searching for different forms of the auxiliary *have* (e.g., *'ve*, *has*, *'s*) and then filtering hits according to the PP leveling/present perfect distinction. Geeraert and

Newman's (2011) study, however, uses pre-made verb combinations prepared by the authors based on the verbs found in Geeraert's (2010) thesis. Such a method may have missed some PP-leveling occurrences if the verbs in these occurrences were not predicted by the authors. Nevertheless, Geeraert and Newman's (2011) study examined PP leveling in both present and past perfect constructions, which may have yielded higher numbers of leveled constructions. To ascertain the reasons for discrepancies in the overall frequency of PP-leveling constructions, further quantitative corpus-based studies are required.

Comparing the frequency of PP leveling in SBCSAE with American corpora in Chatten et al. (2024) reveals a somewhat contrasting pattern. Chatten et al. (2024) report that ~3% of present perfect constructions were leveled in Switchboard and ~15% in PNC. Authors do not explicitly suggest any reasons for the drastic difference in leveling rates between the Switchboard and PNC, but they do indicate that the difference may be due to the corpora's demographics and design. As SBCSAE exhibits only ~1% of leveling, it is much closer in leveling rate to Switchboard than PNC. As Switchboard and SBCSAE are much more similar in the type of speech they consist of, one obvious suggestion would be that leveling rates are lower in more spontaneous conversations. This idea, however, would go against the proposal, made multiple times in this thesis and earlier research, that leveled forms become more leveled as the style of the conversation becomes more casual and spontaneous.

This apparent contradiction suggests that PP leveling is a complex phenomenon that is not determined by a single factor. Such a notion would also uphold the continuum hypothesis, as PP leveling appears to be both stable and influenced by multiple complex continuous factors, rather than categorical factors, such as casual/careful style or spontaneous/elicited speech. Another possible explanation for the similarity in frequency between Switchboard and SBCSAE is the time frame of the corpora compilation. Recordings for both corpora were made in the 1990s, while the last recordings for PNC were made in 2010. It may simply be that the

leveled forms have become more frequent over the past 20 years. To test that, one would need to conduct a diachronic corpus study with a more recent dataset, e.g., a corpus collected in the 2020s. And lastly, it is also possible that the differences in frequency are connected to the small size of SBSCAE, with only 9 leveled constructions identified. All in all, due to the limited size of the dataset in this thesis, any generalizations should be taken with caution.

Intra-speaker variation and self-repair were observed during the analysis, possibly suggesting that both forms are entrenched in some speakers' minds and used interchangeably. Most usage-based theories argue that intra-speaker variation is unlikely to be a performance error and often reflects patterns of linguistic knowledge on the individual level (see 1.4.3). The same can also be argued based on the findings of this thesis. The recurrence of PP leveling within the same recording makes a purely accidental performance error a less plausible explanation, especially given the absence of immediate self-repair and the coexistence of leveled and standard forms in the same speaker's speech. Instead, it may point to a systematic property of the speaker's mental grammar. While one instance of PP leveling could be linked to contextual factors such as modal presence, the recurrence across different contexts strengthens the argument for individual grammatical variation. At the same time, Lajuan consistently produces standard present perfect forms elsewhere, indicating stable intra-speaker variation rather than categorical non-standard usage. An examination of verb classes and frequencies in the speaker's speech provides limited support for a frequency-based explanation: although highly frequent verbs appear resistant to PP leveling due to entrenchment effects, once these are excluded, the remaining verb classes are used in roughly balanced proportions. As such, there is no clear evidence that PP leveling in this speaker is driven by a preference for specific verb classes, suggesting that verb class frequency alone cannot explain this variation.

Previous studies observed that PP leveling frequently co-occurs with modals. Out of nine occurrences of PP leveling in the SBSCAE corpus, three co-occurred with modals. Table

9 shows the verb lemmas and modals with which they were found to co-occur in the SBCSAE. Table 9 also presents the percentage that shows the preference for use of the verb lemma in the past tense form when the construction includes a modal found in Geeraert and Newman's (2011) study¹¹. As shown in Table 9, the percentages for all three verbs are quite similar. The highest percentage in Geeraert and Newman's study was 64% for the verb *freeze*, and the lowest 15% for *drink*. Thus, the three verbs identified in the present study may be characterized as having a medium tendency to co-occur with modals in PP leveling. The reason for the appearance of these verbs rather than the more probable ones in the SBCSAE may again be the size of the corpus or simply the general lower frequency of such verbs as *freeze* in the corpus. In this thesis, it was found that every modal-PP leveling co-occurrence included the modal *would*. It may be the case that *would* co-occurs with PP leveling more often than other modal verbs – this was not investigated in Geeraert and Neman's (2011) study. It is one of the interesting findings that require further quantitative investigation to ascertain any tendencies.

Table 9. Co-occurrence of modals with verb lemmas in SBCSAE, presented with example number and the percent difference in favor of PP leveling co-occurring with modals from Geeraert and Newman (2011).

Example number	Verb lemma	Modal	Percent difference in favor of co-occurrence with modals
(9)	<i>choose</i>	would	39%
(14)	<i>go</i>	would	42%
(24)	<i>take</i>	would	39%

The co-occurrence of modals with PP leveling was discussed in detail in the results section of this thesis. Several possible reasons for such a tendency can be proposed. Firstly, it

¹¹ Geeraert and Newman (2011) do not provide any absolute or normalized numbers for the total number of leveled constructions, so the scale is unknown.

is evident that the meaning of the construction as a whole changes, as the present perfect within the modal construction changes as it becomes non-finite. The present perfect without a modal may express continuation of an action started in the past, as well as a result or an experience in addition to the past. However, during analysis, it was discovered that the non-finite present perfect usually loses these semantic properties. For example, in the sentence “I would’ve taken a shower if I knew I had to go out,” *would’ve taken* expresses a hypothetical situation in the past that did not occur and has no result or experience aspect in its meaning. Within the modal construction, the modal expresses either epistemic or deontic meanings, while the present perfect expresses only the past. Thus, if the co-occurrence with a modal influences the meaning of the present perfect within the modal construction, it may also influence the form in which the present perfect appears within the construction, hence PP leveling. While this thesis initially treats modal perfects as variants of the present perfect, the analysis suggests they may be considered a distinct construction. For comparability with earlier work, they are nevertheless included under the umbrella of PP leveling. Figure 7 demonstrates the competing standard present perfect and PP leveling modal constructions, their forms, and meanings.

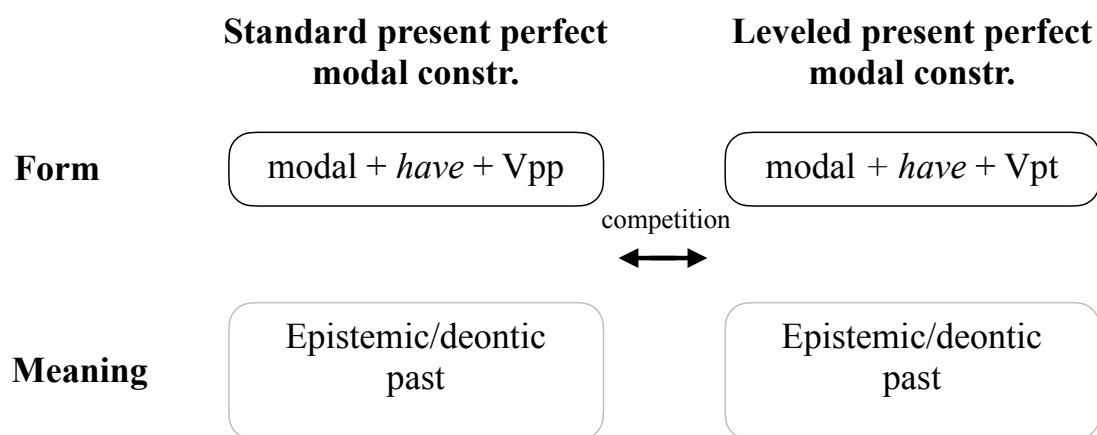


Figure 7. Schematic representation of the forms and meanings of the competing standard present perfect and PP-leveled modal constructions.

Previous research has also shown that the subject person influences leveling. First-person subjects were reported to induce leveling, whereas third-person subjects were reported to inhibit it. Table 10 shows the person of the subject that appears with each leveled construction, as well as the form of the subject. In total, the subject is in the first person 6 times and in the third person 3 times. Such a distribution does not contradict previous findings that indicate a preference for the leveled form when the subject is in the first person (Chatten et al 2024) but they also do not provide independent confirmation of them. The verb *choose* appears all three times with the first-person subject, which can indicate a possible stronger influence on the leveling of that verb by the first-person subject. As this may also be a coincidence due to the limited number of leveled tokens in the SBSCAE, this hypothesis should be tested in future studies.

Table 10. Subject person by example number with verb lemmas and the form of the subject.

example #	verb lemma	subject person	subject
(9)	choose	first	<i>I</i>
(12)	choose	first	<i>I</i>
(13)	choose	first	<i>I</i>
(14)	go	third	<i>he</i>
(16)	go	first	<i>I</i>
(18)	see	third	<i>God</i>
(19)	beat	third	<i>Alisheba</i> (proper noun)
(21)	speak	first	<i>I</i>
(24)	take	first	<i>I</i>

Overall, the findings indicate that past participle leveling in the present perfect occurs infrequently in present-day spoken American English. Given the contradictory frequency estimates reported in previous studies, establishing realistic expectations was challenging. Nevertheless, it was expected that at least 3% (as reported in Chatten et al. 2024) of present perfect constructions would be leveled in the SBCSAE corpus. This expectation was not met, most likely due to the limited size of the corpus. A finding of previous research, the inducing influence of the first-person subjects on leveling, was also illustrated by examples in this thesis, as leveled forms tended to co-occur with the first-person subjects. One verb, *choose*, always co-occurred with a first-person subject in the SBCSAE in its leveled form, although the absolute number of tokens was only 3. The verb *choose* should be examined separately in a future quantitative study to determine whether a correlation exists. The previously reported tendency of leveled forms to co-occur with modals (Geeraert & Newman 2011; Chatten et al 2024) found mixed evidence in this thesis, as only 3 out of 9 leveled constructions co-occurred with modals in the SBCSAE. It was also found that modal perfects have different meanings, and therefore it is proposed that further research should examine them as constructions separate from the finite perfect. In all 3 modal perfect constructions, the modal was *would*. This suggests a possible

tendency for *would* to be especially inducing for the PP leveling. This tendency should also be studied further.

Future research should examine PP leveling in larger, more balanced corpora that include a wider range of contexts, which may yield a greater number and greater diversity of occurrences. Importantly, future investigations should continue to complement formal or elicited data with spontaneous speech. Additionally, a research gap remains in experimental approaches to leveled forms; this could be addressed through methods such as acceptability judgment tasks. One potential research question is whether the use of PP-leveled constructions in certain linguistic contexts (e.g., first-person subject pronouns and modal verbs) is judged more acceptable than others. A comparative learner-English approach is also a promising avenue for further inquiry, as the frequency and contexts of leveled forms can be compared between native and learner English. The observed intra-speaker variation provides another potential topic for future investigation. A qualitative study based on a larger corpus and focusing on individual speakers could provide deeper insight into this phenomenon.

CONCLUSION

This thesis addressed the phenomenon of past participle leveling in spoken American English. Possibly due to its low frequency and uncertainties in terminology, there has not been sufficient research on the phenomenon in recent years. Additionally, research examining the phenomenon qualitatively has been especially scarce. By drawing on spontaneous spoken data and analyzing it within the Construction Grammar framework, this study aimed to contribute a qualitative perspective on an area that has often been addressed primarily through quantitative corpus analyses. The main aim of the thesis was to investigate how past participle leveling compares to the standard present perfect with respect to form, meaning, and usage context in spoken American English. The study discussed these aspects of the leveled construction based on data from the Santa Barbara Corpus of Spoken American English.

The thesis was structured as follows. The first chapter introduced the phenomenon of PP leveling and situated it within previous research on the English verbal paradigm and present perfect constructions. It outlined the relevant grammatical background, discussed earlier findings on frequency and distribution, and introduced Construction Grammar as the theoretical framework adopted in the study. This framework was selected for its suitability for analyzing form–meaning pairings and variation in competing constructions. The second main chapter described the data and methodology used in the empirical analysis. It introduced the corpus, explained the selection and annotation of relevant constructions, and outlined the qualitative analysis procedure. This chapter also clarified how instances of the standard present perfect and leveled variants were identified and categorized, providing the methodological basis for the analysis.

Reviewing previous research revealed that PP leveling in present perfect constructions has been examined primarily in terms of irregular verb morphology, verbal paradigm, and language change, often with a quantitative focus. Previous research has consistently shown that

leveling is most likely to occur with certain classes of irregular verbs and that it remains a relatively infrequent phenomenon in contemporary English. However, earlier studies differ in their interpretation of its grammatical status, with some treating it as a performance-based variation and others as a stable non-standard pattern. In previous examinations, relatively little attention has been paid to the phenomenon from a qualitative or construction-based perspective, especially using spontaneous spoken data. These observations provided the theoretical and empirical motivation for the present study and informed the methodological choices made in the thesis.

The analysis showed that past participle leveling is attested in the corpus, but it occurs much less frequently than standard present perfect constructions. The data also suggested that PP leveling is not random. Leveling occurs more frequently with specific classes of irregular verbs and follows recognizable morphological patterns. The analysis did not reveal a prominent semantic difference between leveled and standard forms, as both were shown to express similar aspectual meanings in context. Only the modal perfect construction was found to differ in meaning (more on that later). In terms of form, the data may indicate that two formally distinct, yet functionally similar constructions coexist and compete in English. Moreover, the examination of usage contexts indicated that PP leveling is not restricted to a single register or type of conversation within spoken discourse. A key finding is the intra-speaker variation observed in one speaker in the SBCSAE. This finding may indicate that speakers can alternate between standard and leveled forms, and that instances of self-repair can result from this alternation.

In summary, the study's results provide a coherent answer to the research question. PP leveling in the present perfect emerges as a marginal variant in spoken American English. The analysis suggests that this variation can be accounted for in the CxG framework as competition between two constructions that are similar in meaning and differ mainly in morphological form.

These conclusions, however, are limited by the corpus size and nature, as well as the qualitative approach to the analysis. Moreover, the findings are representative of the spoken language in the SBCSAE corpus and may not reflect overall trends in English. Overall, the thesis contributes to existing research by offering a detailed qualitative account of a low-frequency but theoretically relevant phenomenon. By showing how past participle leveling can be characterized across form, meaning, and usage, the study shows that paying attention to subtle, detailed differences is important when analyzing English grammar and supports the value of a construction-based approach to morphosyntactic variation.

REFERENCES

- Anderwald, Lieselotte. 2009. *The Morphology of English Dialects*. Cambridge: Cambridge University Press.
- Anderwald, Lieselotte. 2012. Variable past-tense forms in nineteenth-century American English: Linking normative grammars and language change. *American Speech*, 87:3, 257–293.
- Anderwald, Lieselotte. 2020. American English gotten: Historical retention, change from below, or something else? In Ewa Jonsson and Tove Larsson (eds). *Voices Past and Present - Studies of Involved, Speech-related and Spoken Texts*, 187–204. Amsterdam and Philadelphia: John Benjamins.
- Atwood, Bagby E. 1953. *Survey of Verb Forms in the Eastern United States*. Ann Arbor: University of Michigan Press.
- Bülow, Lars, Simone E. Pfenninger, Markus Schiegg, and Alexander Werth. 2021. Introduction. In Alexander Werth, Lars Bülow, Simone E. Pfenninger, Markus Schiegg (eds.). *Intra-Individual Variation in Language*, 1–8. Berlin, Boston: Walter de Gruyter Mouton.
- Bybee, Joan and Carol Lynn Moder. 1983. Morphological classes as natural categories. *Language*, 59:2, 251–270.
- Bybee, Joan. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.
- Bybee, Joan. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes*, 10:5, 425–455.
- Bybee, Joan. 2006a. *Frequency of Use and the Organization of Language*. Oxford, New York: Oxford University Press.
- Bybee, Joan. 2006b. From Usage to Grammar: The Mind's Response to Repetition. *Language*, 82: 4, 711–733.
- Campbell-Kibler, Kathryn. 2007. Accent, (ING), and the Social Logic of Listener Perceptions. *American Speech*, 82: 1, 32–64.
- Capelle, Bert. 2006. Particle placement and the case for “allostructions”. *Constructions*, SVI–7, 1–28.
- Carstairs, Andrew. 1983. Paradigm economy. *Journal of Linguistics*, 19:1, 115–128.
- Chambers, Jack K. 2013. Patterns of Variation Including Change. In Jack K. Chambers and Natalie Schilling (eds.). *The Handbook of Language Variation and Change: Second Edition*. Chichester: John Wiley and Sons, Inc.

- Chatten, Alicia, Kimberley Baxter, Erwanne Mas, Jailyn Peña, Guy Tabachnick, Daniel Duncan, Laurel MacKenzie. 2024. "I've Always Spoke Like This, You See": Preterite-to Participle Leveling in American and British Englishes. *American Speech*, 99:1, 3–46.
- Cognola, Federica, Ivano Baronchelli and Evelina Molinari. 2019. Inter- vs. Intra-Speaker Variation in Mixed Heritage Syntax: A Statistical Analysis. *Frontiers in Psychology*, 10:1528, 1–20.
- Croft, William and Alan D. Cruse. 2004. *Cognitive Linguistics*. Cambridge; New York; Melbourne; Madrid; Cape Town; Singapore; São Paulo: Cambridge University Press.
- Davies, Mark. 2008. *The Corpus of Contemporary American English (COCA)*. Available at <https://www.english-corpora.org/coca/>, accessed March 20, 2026.
- Du Bois, John W., Wallace L. Chafe, Charles Meyer, Sandra A. Thompson, Robert Englebretson, and Nii Martey. 2000–2005. *Santa Barbara Corpus of Spoken American English, Parts 1–4*. Philadelphia: Linguistic Data Consortium.
- Eiskovits, Edina. 1987. Variation in the lexical verb in inner-Sydney English. *Australian Journal of Linguistics*, 7:1, 1–24.
- Ellsäßer, Sophie and Oliver Schallert. 2021. Intra-individual Variation in Morphosyntax: A Constraint-based Perspective. In Alexander Werth, Lars Bülow, Simone E. Pfenninger, Markus Schiegg (eds.). *Intra-Individual Variation in Language*, 207–242. Berlin, Boston: Walter de Gruyter Mouton.
- Feagin, Crawford. 1979. Variation and Change in Alabama English: A Sociolinguistic Study of the White Community. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 26: 2, 250–251.
- Fowler, Henry W. and Robert Burchfield. 1998. *Fowler's Modern English Usage: Revised Third Edition*. Oxford and New York: Oxford University Press.
- Gardiner, Shayna and Naomi Nagy. 2017. Stable Variation vs. Language Change and the Factors that Constrain Them. *University of Pennsylvania Working Papers in Linguistics: Selected Papers from NWAV 45*, 23:2, 78–85.
- Geeraert, Kristina and John Newman. 2011. I haven't drank in weeks: The use of past tense forms as past participles in English corpora. *Language and Computers*, 73: December, 13–33.
- Geeraert, Kristina. 2010. I haven't drank in weeks: preterite shift in English. Unpublished MA thesis. University of Alberta, Edmonton, Canada.
- Goldberg, Adele E. 1995. *Constructions: A Constructional Approach to Argument Structure*. Chicago: The University of Chicago Press.

- Goldberg, Adele E. 2013. Constructionist approaches. In Thomas Hoffmann and Graeme Trousdale (eds). *The Oxford Handbook of Construction Grammar*, 14–31. New York: Oxford University Press.
- Goldberg, Adele E. 2019. *Explain Me This: Creativity, Competition, and the Partial Productivity of Constructions*. Princeton and Oxford: Princeton University Press.
- Hernández-Campoy, Juan Manuel and Juan Antonio Cutillas Espinosa. 2013. The effects of public and individual language attitudes on intra-speaker variation: A case study of style-shifting. *Multilingua: Journal of Cross-Cultural and Interlanguage Communication*, 32:1, 79-101.
- Hilpert, Martin. 2013. *Constructional Change in English: Developments in Allomorphy, Word Formation, and Syntax*. Cambridge; New York; Melbourne; Madrid; Cape Town; Singapore; São Paulo; Delhi; Mexico City: Cambridge University Press.
- Hilpert, Martin. 2025. Frequency: Psychological and Methodological Considerations. In Miriam Fried and Kiki Nikiforidou (eds.). *The Cambridge Handbook of Construction Grammar*, 149–170. Cambridge: Cambridge University Press.
- Hoffman, Thomas, and Graeme Trousdale. 2011. Variation, Change and Constructions in English. *Cognitive Linguistics*, 22:1, 1–23.
- Hoffmann, Thomas, and Mark Turner. 2025. Creative Construction Grammar. In Sarah Duffy and Nick Riches (eds.). *Elements in Cognitive Linguistics*. Cambridge: Cambridge University Press.
- Hudson, Richard A. 2007. English Dialect Syntax in Word Grammar. *English Language and Linguistics*, 11:2, 383–405.
- Hundt, Marianne. 2009. Colonial Lag, Colonial Innovation or Simply Language Change? In Günter Rohdenburg and Julia Schlüter (eds.). *One Language, Two Grammars?: Differences between British and American English*, 13–37. Cambridge: Cambridge University Press.
- Iwata, Seizi. 2014. Construction grammar. In Andrew Carnie, Yosuke Sato and Daniel Siddiqi (eds). *The Routledge Handbook of Syntax*, 647–669. London and New York: Routledge.
- Klavan, Jane. 2021. The Alternation Between Exterior Locative Cases and Postpositions in Estonian Web Texts. *Eesti Ja Soome-Ugri Keeleteaduse Ajakiri*, 12:1, 153-88.
- Labov, William. 1966. *The social stratification of English in New York City*. Cambridge: Cambridge University Press.

- Labov, William. 1984. Field methods of the project on linguistic change and variation. In John Baugh and Joel Sherzer (eds.). *Language in use: readings in sociolinguistics*, 28–54, Englewood Cliffs: Prentice-Hall.
- Labov, William. 2001. *Principles of linguistic change: Vol. II, Social factors*. Malden and Oxford: Blackwell Publishers.
- Lass, Roger. 1994. Proliferation and option-cutting: The strong verb in the fifteenth to eighteenth centuries. In D. Stein and I.T.-B. Ostade (eds). *Towards a Standard English: 1600–1800*, 81–113. Berlin and New York: Mouton de Gruyter.
- Lass, Roger. 1999. Phonology and morphology. In Roger Lass (ed). *The Cambridge History of the English Language, Volume III: 1476–1776*, 56–186. Cambridge: Cambridge University Press.
- MacWhinney, Brian. 2000. The CHILDES Project: Tools for Analyzing Talk (3rd ed.). *Computational Linguistics*, 26:4, 657–657.
- Molenaar, Peter C. M. 2008. On the implications of the classical ergodic theorems: analysis of developmental processes has to focus on intra-individual variation. *Developmental psychobiology*, 50: 1, 60-9.
- Nickel, Grit. 2021. Intra-individual Variation in Nominal Inflection: Analyses of Directly Elicited Data of the Bavarian Linguistic Atlas. In Alexander Werth, Lars Bülow, Simone E. Pfenninger, Markus Schiegg (eds.). *Intra-Individual Variation in Language*, 177–206. Berlin, Boston: Walter de Gruyter Mouton.
- OED Online. Oxford English Dictionary Online. 2024. Available at: <http://doi.org/10.1093/OED/5588938349>, accessed March 21, 2026.
- Peters, Pam. 2009. Irregular verbs: Regularization and ongoing variability. In Pam Peters, Peter Collins, and Adam Smith (eds). *Comparative Studies in Australian and New Zealand English: Grammar and Beyond*, 13–30. Amsterdam: John Benjamins Publishing Company.
- Põldvere, Nele and Carita Paradis. 2019. ‘What and then a little robot brings it to you?’ The reactive what-x construction in spoken dialogue. *English Language and Linguistics*, 24:2, 307–332.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London and New York: Longman.
- Roomäe, Kärt. 2021. *How come that’s a construction? A construction grammar approach to the how come construction in spoken American English*. Unpublished MA thesis. Department of English, University of Tartu, Tartu, Estonia.

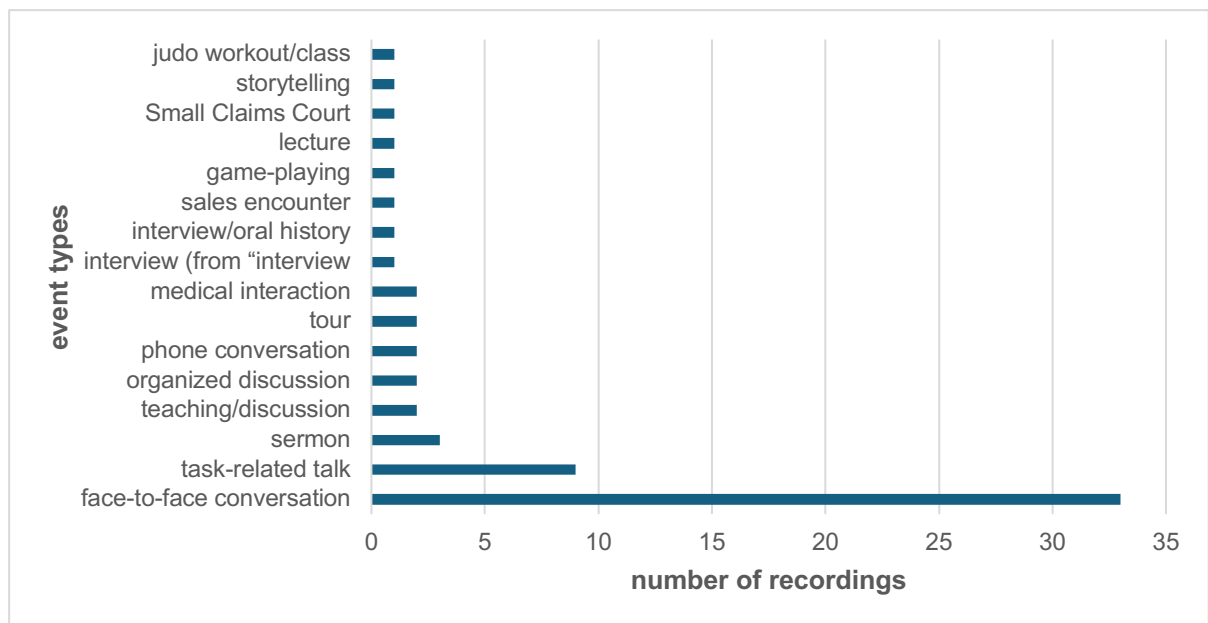
- Schilling, Natalie. 2013. Investigating Stylistic Variation: Second Edition. In J. K. Chambers and Natalie Schilling (eds.). *The Handbook of Language Variation and Change*, 327–249. Chichester: Wiley–Blackwell.
- Shnukal, Anna. 1978. A sociolinguistic study of Australian English: Phonological and syntactic variation in Cessnock. *Dissertation Abstracts International*, 39:12, 7324.
- Taylor, Ann. 1994. Variation in past tense formation in the history of English. *Penn Working Papers in Linguistics*, 1, 143–158.
- Uhrig, Peter. 2015. Why the Principle of No Synonymy is Overrated. *Zeitschrift für Anglistik und Amerikanistik*, 63: 3, 323–337.
- Ulbrich, Christiane and Alexander Werth. 2021. What Is Intra-individual Variation in Language? In Alexander Werth, Lars Bülow, Simone E. Pfenninger, Markus Schiegg (eds.). *Intra-Individual Variation in Language*, 9–43. Berlin, Boston: Walter de Gruyter Mouton.
- Wolfram, Walt and Donna Christian. 1975. *Sociolinguistic Variables in Appalachian Dialects*. Washington, D.C.: Center for Applied Linguistics.

Appendix 1: Types of Recordings in SBCSAE

The list of files/recordings in the SBCSAE, including number of speakers, event types, and short event descriptions. Files sbc001–sbc030 and sbc047–sbc060 were adapted from the metadata (Du Bois et al 2000–2005). However, due to the absence of some metadata files for part III of the corpus, information for files sbc031–sbc046 was adapted from the descriptions of the recordings at the beginning of each file (see Appendix 2). A summarizing graph is available at the very end of the Appendix. The total number of recordings shown in the graph exceeds the total of 60 recordings in the corpus. This is due to recordings being counted separately in each category when they are attributed to two or more categories (e.g., *task-related talk/face-to-face conversation* as in sbc009).

File	Num Speakers	Event Type	Event
sbc001	3	face-to-face conversation	mother and daughter catching up after daughter's absence
sbc002	4	face-to-face conversation	After dinner conversation
sbc003	3	face-to-face conversation	a conversation between friends while preparing dinner
sbc004	6	face-to-face conversation	drinking coffee
sbc005	2	face-to-face conversation	chatting in bed
sbc006	2	face-to-face conversation	2 participants hadn't seen each other for several weeks
sbc007	2	face-to-face conversation	sisters talking about problems, friend in hospital, kid
sbc008	4	interview/task-related talk	trial preparation
sbc009	2	task-related talk/face-to-face conversation	girl helps boyfriend with math problems
sbc010	3	face-to-face conversation	talking about political problems of an arts organization
sbc011	3	face-to-face conversation	face to face conversation
sbc012	9	teaching/discussion	teacher discussing history of Chicano Studies to students
sbc013	5	face-to-face conversation	dinner party for Kendra's birthday, dinner, gifts, cake and singing
sbc014	5	task-related talk	loan officer and 2 board members having a loan meeting
sbc015	3	face-to-face conversation	friends visiting
sbc016	2	sales encounter	customer getting information from audio salesman on purchasing a cassette deck
sbc017	2	face-to-face conversation	speakers had never met before
sbc018	5	task-related talk	everyday activities in veterinarian office
sbc019	5	face-to-face conversation	family conversing with out-of-town visiting relative
sbc020	2	sermon	sermon during staff meeting for campus advisors
sbc021	1	sermon	Sunday morning meeting of a church
sbc022	3	task-related talk	trainee receives job feedback and evaluation
sbc023	11	organized discussion	women's monthly book club meeting
sbc024	2	game-playing	playing Risk and teaching Hearts
sbc025	1	lecture	adult-ed class on Protestant Reformation debates
sbc026	7+	organized discussion	city meeting about Empowerment funding for community projects
sbc027	1	teaching/discussion	science experiments performed and explained
sbc028	3	phone conversation	phone conversation
sbc029	2	task-related talk	engineer asks about furnace work; customer shows him around
sbc030	1	sermon	pastor giving Sunday morning sermon
sbc031	4	face-to-face conversation	Face-to-face conversation between daughters and mother recorded in a restaurant.
sbc032	6	face-to-face conversation	Conversation between neighbors at an outdoor party

sbc033	8	face-to-face conversation	Family argument/discussion. The participants are all relatives or close friends.
sbc034	2	face-to-face conversation	Late night face to face conversation between a married couple.
sbc035	8	face-to-face conversation	Lively family argument/discussion recorded in the kitchen of a family home
sbc036	3	face-to-face conversation	Talk about a baby, friends and coworkers between siblings and a friend
sbc037	4	task-related talk	A family conversation in the kitchen making tamales. The recording contains English-Spanish codeswitching
sbc038	1	tour	This segment is part of a tour of Hoover Dam. The presentation is highly practiced. The main speaker also answers audience questions.
sbc039	6	task-related talk	a training meeting recorded at an aquarium
sbc040	3	tour	Scripted tour of the Kentucky Horse Park / Museum.
sbc041	2	medical interaction	Medical interaction. A patient is consulting with her dietician regarding management of diabetes.
sbc042	6	task-related talk/face-to-face conversation	a family argument
sbc043	2	face-to-face conversation	talk between mother and daughter about their day and acquaintances
sbc044	2	face-to-face conversation	Two friends talking about family and friends, and their experiences as gay men
sbc045	2	face-to-face conversation	Two friends talking while watching TV
sbc046	3	medical interaction	A patient is consulting with his orthopedist regarding a knee injury from a recent skiing accident.
sbc047	2	face-to-face conversation	Two cousins talking about work and life
sbc048	4	face-to-face conversation	Christmas morning, opening Christmas presents
sbc049	10	face-to-face conversation	Family birthday celebration, talking poolside
sbc050	4	face-to-face conversation	Roommates chatting
sbc051	4	face-to-face conversation	Conversation and dinner
sbc052	4	phone conversation	Calling relatives in San Antonio on Christmas
sbc053	7	Small Claims Court	Cases presented to pro tem judge
sbc054	1	storytelling	Arranged storytelling performance
sbc055	2	interview/oral history	Speaker telling life stories, answering questions
sbc056	6	face-to-face conversation	Bill of sale for a horse; ranch tour
sbc057	6	judo workout/class	Warmups, class, participant honored for brown belt
sbc058	2	face-to-face conversation	Mother and son talking while cooking dinner
sbc059	4	face-to-face conversation	Visiting family on Christmas Eve
sbc060	2	face-to-face conversation	Talking while taking a break from work



Appendix 2: An Example of a .cha File in SBSCAE

This figure demonstrates the structure and visual representation of the .cha (CHAT) files in the SBSCAE.

```
@UTF8
@Begin
@Languages:      eng
@Participants:   LENO LENORE Speaker, LYNN LYNNE Speaker, DORI DORIS Speaker, ENV ENV Environment
@Options:        CA, bullets
@ID:             eng|SBSCAE|LENO||||Speaker|||
@ID:             eng|SBSCAE|LYNN||||Speaker|||
@ID:             eng|SBSCAE|DORI||||Speaker|||
@ID:             eng|SBSCAE|ENV||||Environment|||
@Media: 01, audio
@Comment:        Actual Blacksmithing.
@Comment:        This is a conversation recorded in
                  rural Hardin, Montana. Mae Lynne is a student of equine science, and
                  is the main speaker. She is telling Lenore (a visitor and near
                  stranger) about her studies. Doris, Mae Lynne's mother, is doing
                  housework, but joins the conversation near the end to discuss
                  friends of their family.
*LENO:           (..) So you don't need to go (..) borrow equipment from anybody 0_9210
                  to +/. 9210_9520
*LENO:           (..) to do the feet ? 9520_14100
*LENO:           (..) [ Do the hooves ] ? 14100_15780
*LYNN:           [ &=in &=lengthened ] YWell 15010_16780
                  we're gonna have to find somewhere 16780_18320
                  to get 18330_18850
                  &=ex (..) somethingY &=ex . 18850_20690
*DORI:           (..) So 20690_21190
                  [ Mae- ] +... 21190_21740
*LYNN:           [ I'm gonna ] &=ex +... 21260_22240
*DORI:           [Mae Lynne XX ] . 22240_23230
*LYNN:           [&=in We're not] gonna do the feet today 22280_24250
                  I'm gonna wait till like 24250_25070
                  early in the morning: 25070_26140
                  (..) to do those 26140_26620
                  cause y- +/. 26620_26870
*LYNN:           I mean you get s:o ti:red . 26870_28370
*LYNN:           &=in (..) n- you just 28370_30100
                  (..) it takes ?uh +/. 30100_32360
*LYNN:           well 32360_32590
                  it takes me longer than most people 32590_33900
                  cause you know 33900_34220
                  I'm not as stro:ng and 34220_35230
                  &=in (..) and I'm not as good 35230_36880
                  as like somebody that would do it (..) all the ti:me . 36880_39300
*LYNN:           (..) You know . 39300_39800
*LYNN:           (..) I mean 39800_40270
```

Appendix 3: List of Standard Irregular Verb Forms

This appendix presents a list of irregular verb forms adapted from Quirk et al. (1985: 115–120) considered standard in this study.

<i>bear-bore-borne</i>	<i>rise-rose-risen</i>
<i>beat-beat-beaten</i>	<i>run-ran-run</i>
<i>begin-began-begun</i>	<i>see-saw-seen</i>
<i>bid-bid/bade-bidden</i>	<i>shake-shook-shaken</i>
<i>bite-bit-bitten</i>	<i>show-showed-shown</i>
<i>blow-blew-blown</i>	<i>shrink-shrank-shrunk</i>
<i>break-broke-broken</i>	<i>sing-sang-sung</i>
<i>choose-chose-chosen</i>	<i>sink-sank-sunk</i>
<i>come-came-come</i>	<i>slay-slew-slain</i>
<i>draw-drew-drawn</i>	<i>speak-spoke-spoken</i>
<i>drink-drank-drunk</i>	<i>spring-sprang-sprung</i>
<i>drive-drove-driven</i>	<i>steal-stole-stolen</i>
<i>eat-ate-eaten</i>	<i>stride-strode-stridden</i>
<i>fall-fell-fallen</i>	<i>swear-swore-sworn</i>
<i>fly-flew-flown</i>	<i>swim-swam-swum</i>
<i>freeze-froze-frozen</i>	<i>take-took-taken</i>
<i>give-gave-given</i>	<i>tear-tore-torn</i>
<i>go-went-gone</i>	<i>throw-threw-thrown</i>
<i>grow-grew-grown</i>	<i>tread-trod-trodden</i>
<i>hide-hid-hidden</i>	<i>wake-woke-woken</i>
<i>know-knew-known</i>	<i>wear-wore-worn</i>
<i>lie-lay-lain</i>	<i>weave-wove-woven</i>
<i>ride-rode-ridden</i>	<i>write-wrote-written</i>
<i>ring-rang-rung</i>	

Appendix 4: CLAN Commands Used to Search the SBSCAE

This appendix presents the list of commands created by the author in the CLAN analysis program to identify constructions that include the verb *have*.

1. The combination of “’ve” with anything else after it in all .cha extension files:

```
combo +s"\'e^" +w1 *.cha
```

2. The combination of “has” with anything else after it in all .cha extension files:

```
combo +s"has^*" +w1 *.cha
```

3. The combination of “hasn’t” with anything else after it in all .cha extension files:

```
combo +s"hasn't^*" +w1 *.cha
```

4. The combination of “have” and anything else after it in all .cha extension files:

```
combo +s"have^*" +w1 *.cha
```

5. The combination of “haven’t” and anything else after it in all .cha extension files:

```
combo +s"haven't^*" +w1 *.cha
```

6. The combination of “’s” and anything else after it in all .cha extension files:

```
combo +s"*'s^*" +w1 *.cha
```

RESÜMEE

TARTU ÜLIKOOL
ANGLISTIKA OSAKOND

Daria Zimarin

Past Participle Leveling in the Present Perfect Construction: a Corpus-Based Qualitative Analysis

Mineviku partitsiibi nihe inglise oleviku perfekti konstruktsioonis: korpusepõhine kvalitatiivne uurimus

Magistritöö

2026

Lehekülgede arv: 83

Annotatsioon:

Magistritöö eesmärk on uurida mineviku partitsiibi nihet oleviku perfekti konstruktsioonis suulises Ameerika inglise keeles ning selgitada nähtuse olemust konstruktsioonigrammatika raamistikus. Töö koosneb kahest põhiosast: teoreetilisest osast, mis käsitleb inglise keele verbisüsteemi, varasemaid uuringuid ja konstruktsioonigrammatikat, ning empiirilistest osast, mis keskendub korpusanalüüsile.

Uuring põhineb Santa Barbara suulise Ameerika inglise keele korpuse (Santa Barbara Corpus of Spoken American English) kvalitatiivsel analüüsil. Korpusest tuvastati nii normatiivsed kui ka nihkes olevad oleviku perfekti konstruktsioonid, mida analüüsiti süntaktiliste, semantiliste ja konteksti-põhiste tunnuste põhjal.

Analüüs hõlmab üheksat nihkes oleviku perfekti juhtumit, mida võrreldakse normatiivsete konstruktsioonidega. Tähelepanu pööratakse ka *have got* ja *have gotten* konstruktsioonidele ning kõneleja-sisesele variatsioonile.

Tulemused näitavad, et nihe on haruldane, kuid süsteemne nähtus, mis esineb peamiselt spontaanses suulises kõnes ning teatud ebaregulaarsete verbidega. Semantilisi erinevusi normatiivse ja nihkes konstruktsiooni vahel üldjuhul ei ilmne.

Märksõnad: Ameerika inglise keel, konstruktsioonigrammatika, kvalitatiivne uuring, korpusepõhine uuring, oleviku perfect

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Daria Zimarin

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

Mineviku partitsiibi nihe inglise oleviku perfekti konstruksioonis: korpusepõhine kvalitatiivne uurimus (Past Participle Leveling in the Present Perfect Construction: a Corpus-Based Qualitative Analysis),

mille juhendaja on Jane Padrik,

1.1.reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2.üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.

3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Daria Zimarin

Tartus, 12.05.2026

Autorsuse kinnitus

Kinnitan, et olen koostanud käesoleva magistritöö ise ning toonud korrektselt välja teiste autorite panuse. Töö on koostatud lähtudes Tartu Ülikooli maailma keelte ja kultuuride instituudi anglistika osakonna oma õppekava magistritöö juhendist ning on kooskõlas heade akadeemiliste tavadega.

Daria Zimarin

Tartus, 12.05.2026

Lõputöö on lubatud kaitsmisele.

Jane Padrik

Tartus, 12.05.2026