Tampere University

Roosamari Kuttila

"YOU MUST HAVE DRANK TOO MUCH"
A Corpus-based Study on the Use of Past Simple Forms
as Past Participles on the Internet

ABSTRACT

This thesis examines a phenomenon called past tense shifting (PTS) in which the past simple form of an irregular verb is used as a past participle. Six irregular verbs were chosen for this study from four different irregular verb classes: *show*, *swell*, *break*, *wake*, *drink* and *go*. Past participle uses of these verbs in the web-based iWeb corpus are examined to determine how frequently the past simple form is used in place of the standard past participle form.

The past participle is used in three grammatical constructions: the perfect mood, the passive voice, and participial adjectives. The perfect mood can be combined with either the present or past tense to form the present and past perfect and passive constructions can be formed with either the auxiliary *be* or *get*. Furthermore, both present perfect and passive constructions can occur with modal auxiliaries. Therefore, there are seven constructions of interest in this thesis: the present perfect, the past perfect, the modal perfect, the *be*-passive, the *get*-passive, the modal passive, and participial adjectives. The PTS form and standard past participle form of each chosen verb are examined in all of these constructions to determine how frequently each form occurs in each construction. In addition, further observations are made on the use of each form in each of these constructions by examining a sample of 100 tokens.

One of the possible theories to explain past tense shifting is natural morphology. The main thesis of natural morphology is that some features of language are more "natural" than others and language as a whole tends to strive toward more naturalness. Non-standard varieties are generally found to use more natural features than standard varieties and language change frequently happens in the direction of more naturalness. There are two schools of thought in natural morphology: system-independent natural morphology and system-dependent natural morphology. The system-independent approach defines universally natural features which affect all languages; however this theory is found to have little to do with past tense shifting. System-dependent natural morphology, on the contrary, may partially explain this linguistic phenomenon. This version of natural morphology focuses on natural features and patterns that are language-specific, meaning that naturalness is not a universal concept but rather unique to each individual language. In English, a system-defining feature of verbs is an inflectional paradigm where the present simple form is distinguished from the past simple and past participle forms, which are identical to one another. All regular verbs and most irregular verbs follow this pattern, making it the most natural paradigm in English. Therefore, irregular verbs which have distinct past simple and past participle forms may be likely to level the distinction between these two forms to fit the most natural pattern.

In the corpus data certain patterns emerge concerning the use of PTS forms. Firstly, verbs with higher frequency appear to be more resistant to past tense shifting while this phenomenon is more prevalent in lower frequency verbs. Secondly, PTS forms are less frequent in adjectival than verbal uses of the past participle. Thirdly, PTS forms are the most frequent in perfect constructions, particularly modal perfects, which is in line with previous research on past tense shifting. There also seem to be some semantic preferences as regards the PTS and standard form. If a verb has multiple possible meanings, quite often each form is used more frequently with one specific meaning. It is noted that some of these patterns may be related to the higher frequency of certain constructions and meanings in certain registers. As the iWeb corpus is compiled from websites all over the internet, all internet registers are represented. As past tense shifting is a phenomenon presumed to mostly occur in informal registers, PTS form use in informal web registers may be a fruitful topic of further study.

Keywords: Irregular verbs, past tense shifting, past simple, past participle, corpus linguistics

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

TIIVISTELMÄ

Tämä tutkielma tarkastelee ilmiötä, jossa epäsäännöllisen verbin imperfektimuotoa käytetään menneen ajan partisiipin sijaan (eng. past tense shifting eli PTS). Tutkittavaksi on valittu kuusi epäsäännöllistä verbiä neljästä eri epäsäännöllisten verbien luokasta: *show*, *swell*, *break*, *wake*, *drink* ja *go*. Näiden verbien käyttöä menneen ajan partisiippeina tarkastellaan verkkopohjaisessa iWeb -korpuksessa, jotta saadaan selville, kuinka yleistä imperfektimuodon käyttö standardin partisiippimuodon sijaan on.

Menneen ajan partisiippia käytetään *have* -apuverbin kanssa muodostamaan perfekti ja pluskvamperfekti sekä *be*- tai *get* -apuverbin kanssa muodostamaan passiivi. Sekä perfekti että passiivi voivat esiintyä myös modaaliapuverbin kanssa. Lisäksi menneen ajan partisiippi voi toimia adjektiivina. Tässä tutkielmassa tarkastellaan siis seitsemää eri kieliopillista rakennetta: perfekti, pluskvamperfekti, modaalinen perfekti, *be* -passiivi, *get* -passiivi, modaalinen passiivi, sekä adjektiivina toimiva partisiippi. Tutkimuksen tarkoituksena on selvittää, kuinka usein PTS-muotoa ja standardimuotoa käytetään kussakin rakenteessa. Kummankin muodon käytöstä näissä rakenteissa tehdään myös tarkempia huomioita tarkastelemalla 100 esiintymän näytettä.

Eräs teoria, joka mahdollisesti selittää menneen ajan partisiippien korvaamista imperfektimuodoilla on luonnollinen morfologia (eng. natural morphology). Luonnollisen morfologian pääteesi on, että tietyt kielen piirteet ovat luonnollisempia kuin toiset ja kieli yleensä pyrkii kohti maksimaalista luonnollisuutta. Epästandardeissa kielimuodoissa näitä luonnollisia piirteitä usein esiintyy enemmän kuin standardeissa kielimuodoissa. Luonnollisessa morfologiassa on kaksi ajatussuuntaa: järjestelmästä riippumaton luonnollinen morfologia (system-independent natural morphology) sekä järjestelmäsidonnainen luonnollinen morfologia (system-dependent natural morphology). Järestelmästä riippumaton ajatussuunta määrittelee tiettyjä universaalisti luonnollisia kielen piirteitä; tällä ei kuitenkaan ole havaittu olevan tekemistä tutkimuksen alla olevan ilmiön kanssa, toisin kuin järjestelmäsidonnaisella luonnollisella morfologialla. Tämän ajatussuunnan mukaan luonnollisuus ei ole universaali konsepti vaan kullakin kielellä on omat luonnolliset piirteensä. Englannin kielessä eräs näistä luonnollisista piirteistä on verbin taivutus, jossa preesensmuoto on erilainen kuin imperfektimuoto ja menneen ajan partisiippi, jotka ovat puolestaan keskenään identtisiä. Kaikki säännölliset verbit ja suurin osa epäsäännöllisistä verbeistä noudattavat tätä kaavaa. Tämän vuoksi epäsäännölliset verbit, joilla on erillinen imperfekti- ja partisiippimuoto, saattavat yksinkertaistaa taivutustaan noudattaakseen kaikkein luonnollisinta kaavaa.

Korpusdatassa on havaittavissa tiettyjä trendejä PTS-muotojen käytön suhteen. Ensinnäkin korkeamman frekvenssin verbit vaikuttavat käyttävän PTS-muotoja matalan frekvenssin verbejä harvemmin. Toiseksi PTS-muodot ovat harvinaisempia adjektiivina kuin verbinä käytettävien menneen ajan partisiippien kanssa. Kolmanneksi PTS-muotoja käytetään eniten perfektissä ja erityisesti modaalisessa perfektissä, mikä on linjassa aiempien tutkimusten kanssa. PTS-muodolla ja standardilla partisiippimuodolla vaikuttaa myös olevan joitakin semanttisia preferenssejä. Jotkin näistä löydöksistä saattavat liittyä tiettyjen kieliopillisten rakenteiden ja sanojen merkitysten korkeaan esiintyvyyteen tietyissä rekistereissä. Koska iWeb -korpus on koottu eri verkkosivuilta ympäri internetiä, siinä on edustettuna monta eri rekisteriä. Koska imperfektimuodon käyttö partisiippimuodon sijaan on oletettavasti yleisempää epämuodollisissa konteksteissa, olisi tätä ilmiötä mahdollisesti tarpeen tutkia eri rekistereiden välillä.

Avainsanat: epäsäännöllinen verbi, imperfekti, menneen ajan partisiippi, korpustutkimus

Table of Contents

1. Introduction

In the English language, verbs have five distinct forms: the present tense base form, the present tense third person singular, the present participle, the past simple and the past participle (Quirk et al. 1985, 96). While the third person singular and the present participle are always marked in the same way, there are some so-called irregular verbs whose process of forming the past simple and past participle form differs from that of regular verbs. Contrary to regular verbs, some of these irregular verbs have distinct past simple and past participle forms. Sometimes in non-standard English, one of these forms is generalised to fulfil the role of both. When the past simple form is substituted for the past participle form, this is called past tense shifting (PTS) by Geeraert and Newman (2012). The following tokens in the iWeb corpus are examples of this phenomenon:

(1) Sure, we knew that an emergency could have **came** up that cost more than $1000, but we were willing to take that risk anyway. (iWeb, ourfreakingbudget.com)

(2) Like a firework rocket, his Indian career was **began** with a bang and then flamed out. (iWeb, letsgotribe.com)

Though PTS has seen less research than shifting in the opposite direction, that is, using the past participle form in place of the past simple form, this phenomenon is by no means recent. Past tense shifting seems to have first emerged in late Middle English and increased in use in Early Modern English (Lass 1994, 89).

In earlier grammars and usage guides past tense shifting was seen as a usage problem (Van Ostade 2015, 296). However, as modern grammars tend to be descriptive rather than prescriptive, the attitudes toward past tense shifting and other non-standard features of English are quite different today. Non-standard language is seen more as a subject of research than a problem that needs to be fixed. Biber et al. (2021, 1118) list the generalisation of the past tense form to the different functions of the past participle form among the common morphosyntactic variants in non-standard grammar. The term "non-standard" can be seen as

1

problematic, however, as it presupposes that there is one agreed upon form of English that is standard. As there multiple widely spread regional varieties of English, it may be difficult to decide which features should be considered standard and which non-standard. However, Biber et al. (2021, 18) argue that most differences between regional varieties come down to a small range of spelling and lexical variants and so it can be said that there is, indeed, a widely recognised standard English which is codified in grammars, dictionaries and usage guides. Standard English can be defined as the form of the language accepted by the majority of educated speakers (e.g. Quirk et al 1985, 18-19; Biber et al. 2021, 18; Denison 2008, 533). What is generally meant by "non-standard English" is any language pattern deviating from this agreed upon norm. While standard English is used in most formal written texts, non-standard linguistic features are generally restricted to particular regions or social classes and informal contexts (Biber et al. 2021, 20).

One domain where non-standard English is common is the internet. While there are plenty of formal written texts on the internet, a large portion of the language used online is informal (McCulloch 2019, 2-3). McCulloch (ibid.) argues that with the rise of the internet, there are now more informal written texts easily accessible to linguists than ever before. This means that many non-standard features which have previously been mostly restricted to speech can also be found in written language. Some objections have been raised against this written non-standard language as non-standard features tend to be considered less acceptable than standard variants (Baron 2008, 161, Denison 2008, 533). Some go as far as to say that the English language is on the decline because of the internet. Baron (2008, 161-162), however, points out that often what is considered language decline is simply language change in progress. Denison (2008) argues that clues to the direction of language change may be found in non-standard language patterns as new linguistic features often emerge in non-standard language and later spread to standard use. Anderwald (2011a, 185-6) observes that

there does, indeed, seem to be some spreading of non-standard variants at the expense of the standard when it comes to past tense verb forms.

The purpose of this thesis is to study the use of PTS forms on the internet by conducting a corpus-based study utilising the web-based iWeb corpus. In doing this, I aim to find answers to the following research questions:

1. How frequent are the PTS forms of the selected verbs compared to their respective standard forms?

2. In what grammatical contexts do PTS forms most frequently appear?

3. What other relevant observations can be made based on the corpus data?

In order to answer these questions, both quantitative and qualitative methods, which are explained in more detail in section 3.3, are used.

This thesis is divided into four main parts. Firstly, the literary background is introduced in chapter 2 where key grammatical concepts are explained with the help of grammars and the theory of natural morphology is introduced. The findings of earlier studies on past tense shifting are also presented in this chapter. After the theoretical background, the data and methodology used in this study are presented in chapter 3. Chapter 4 shows the results of the corpus study. Finally, chapters 5 and 6 delve into deeper discussion on the findings and deficiencies of this study.

2. Background

In this chapter I outline the theoretical framework of this thesis. In section 2.1 I begin by explaining grammatical concepts central to the research topic of this study. I then go on to shed some light on past tense shifting as a linguistic feature in non-standard English in section 2.2.

2.1 The Past Participle

Participles are verb forms which may function as or like adjectives and can be used in combination with auxiliaries to form different tenses, aspects, moods or voices (Huddleston and Pullum 2002, 78). There are two participle forms in the English language: the present participle, sometimes called the *ing-* participle (e.g. *walking*) and the past participle, sometimes called the *ed-* participle (e.g. *asked*) which is the focal point of this thesis. It is worth noting, however, that despite including the words *past* and *present*, respectively, neither past nor present participle forms are in themselves tensed. Participles are nonfinite verb forms and as such are not marked for number, person, tense or mood (Declerck 1991, 447-48). That being said, participles can be used in combination with other verbs to form compound tenses and moods. In this section I first explain the methods used to form the past participle and then proceed to outline the different constructions in which it is used.

2.1.1 Formation

The past participle form of a regular verb is formed by adding the suffix *-ed* to its base, making it identical to the past simple form (Biber et al. 2021, 392). There are, however, around 200-250 English verbs which do not take the suffix *-ed* but instead form their simple past and past participle forms in a variety of other ways (Quirk et al. 1985, 104). The past participle form of these irregular verbs can be formed by adding a nasal morpheme (e.g.

taken), an alveolar morpheme (e.g. *kept*) or a zero morpheme (e.g. *cut*) or by changing the root vowel (e.g. *swum*) or final consonant (e.g. *built*) (Declerck 1991, 447).

Though it is not of great relevance to this thesis, it should be mentioned that some verbs can appear in both a regular and irregular past participle form. The choice of which of the two forms is used depends on many factors, such as register, grammatical function and regional dialect (notably AmE generally prefers regular forms and BrE prefers irregular forms) (Biber et al. 2021, 396). Nevertheless, it seems that the use of regular past participle forms in favour of irregular ones is becoming more common due to regularisation (ibid.). This use of regular forms removes the need to distinguish between past simple and past participle forms – a concept I will explore in a later section.

Biber et al. (2021, 394-95) divide irregular verbs into seven classes based on the patterns used to form their past simple and past participle forms. The classes and some example verbs can be seen in table 1. For this study I have selected two verbs from classes 3 (*show*, *swell*) and 4 (*break, wake*) and one verb from classes 5 (*drink*) and class 7 (*go*). I have excluded classes 1, 2 and 6 as the past simple and past participle forms in these are identical to one another and thus not suitable for the purposes of this thesis. Having all the relevant irregular verb classes represented in the study ensures results on the use of non-standard participle forms of a wide range of different types of verbs which may be used as a base for further study.

Class	Characteristics	Examples
Class 1	Voiceless suffix *-t* in both past simple and past participle. Many verbs in this class can also take the regular suffix *-ed*.	send, sent, sent build, built, built
Class 2	Change in base vowel and *-t* or *-d* suffix in both past simple and past participle.	keep, kept, kept sell, sold, sold
Class 3	Regular *-ed* suffix in simple past and *-(e)n* or regular *-ed* with a possible change in base vowel in past participle.	show, showed, shown (showed) swell, swelled, swollen (swelled)
Class 4	No suffix in past simple and the suffix *-(e)n* in past participle. Base vowel changes once or twice.	take, took, taken give, gave, given
Class 5	No suffixes, only a change in base vowel. The vowel may change once or twice.	come, came, come drink, drank, drunk
Class 6	No changes to the base form, all three forms are identical.	hit, hit, hit let, let, let
Class 7	One or more completely unrelated forms.	go, went, gone

Table 1. Irregular verb classes based on Biber et al. (2021, 394-95)

2.1.2 The Perfect Aspect

Aspect is a grammatical category related to verbs which reflects how a situation is experienced as regards time (Quirk et al. 1985, 188). The distinction between tense and aspect may not always be clear as they are both related to time in the verb phrase. In fact, Quirk et al. (ibid., 189) argue that the two terms are so closely related in meaning that the distinction between tense and aspect in English grammar is "little more than a terminological convenience". Though both related to time in the verb phrase, tense and aspect have slightly different functions and characteristics. Whereas tense concerns the time orientation of past-present-future and is relative to the time of utterance (TU), aspect describes different time-related features such as the completion or lack of completion of actions, events or states

described by the verb (Biber et al. 2021, 458). There are two aspects in English: progressive and perfect. The progressive aspect describes actions and states which are in progress at the time indicated by the rest of the clause and is formed by combining the auxiliary *be* and present participle (ibid.). The perfect aspect, on the other hand, depicts events or states taking place during a period of time leading up to a specified moment and is marked by the auxiliary verb *have* + past participle (ibid.). Both the perfect and progressive aspect can be combined with the present and past tense to form the present and past progressive and present and past perfect respectively. In addition, the perfect and progressive aspect can be combined with each other to form the perfect progressive (Quirk et al. 1985, 189). Next, I will describe in greater detail the various uses of the perfect aspect which utilises the past participle.

In English, time can be conceived as being divided into two spheres based on relation to TU: the past time-sphere which lies wholly before TU and the present time-sphere which includes the pre-present sector, present sector and post-present sector (Declerck 1991, 86). The present perfect falls into the pre-present sector in which the situation is located before TU but the speaker is more concerned with the state of affairs at TU when reporting it and so it has "current relevance" (ibid.; Quirk et al. 1985, 190). This feature differentiates the present perfect from simple past which describes situations that are wholly in the past and hold no current relevance. Also in contrast to simple past, present perfect locates a situation in a timespan that ranges from some time before TU all the way up to TU whereas simple past designates the action to a specific point in the past (Declerck 1991, 97; Quirk et al. 1985, 190). The choice of which form to use in a sentence therefore depends on whether the described state or action is finished or still ongoing and whether the past time referred to is definite or indefinite.

Quirk et al. (1985, 192) list three cases in which present perfect is always used in favour of simple past:

1)	State leading up to the present:

She **has been** absent for ages.
I **have liked** strawberries my whole life.

2)	Indefinite event(s) in a period leading up to the present:

All my siblings **have had** chicken pox.
Have you **been** to Greece?

3)	Habit in a period leading up to the present:

He **has run** 5km every day since his doctor told him to get more exercise.
I **have suffered** from bad nosebleeds since childhood.

In 1) the contrast to simple past lies in the completeness of the state – simple past implies that the state has already ended (cf. *She was absent for ages – but is now present again*). In 2) present perfect leaves the time of the event undefined while simple past ties it to a specific point in time. Compare: *all my siblings have had chicken pox* (at some unspecified point in time and not necessarily all at once) and *all my siblings had chicken pox* (simultaneously at some specific time). 3) is similar to the habitual past which uses the simple past tense but the difference, similarly to 1), lies in the continuation of the habit up to the present (cf. *He ran 5km every day for two months– but does not do so anymore*).

Huddleston and Pullum (2002, 141-46) distinguish four major uses of the present perfect: the continuative perfect, the experiential perfect, the resultative perfect and the perfect of recent past. These can be seen as the different ways in which the past event can be related to TU. Though the present perfect is often used to describe events that continue from a time in the past up to TU and are still in effect, it can also describe events wholly in the past. In fact, the interpretation which places the event wholly before TU is the default reading in cases where there is no time adjunct such as *since* or *for* to imply continuation (ibid., 141):

[1]

    a. I **have had** long hair since high school.

    b. I **have had** long hair.

These examples illustrate how the absence of a time adjunct leads to a non-continuative

perfect [1b] and the inclusion of the adjunct *since* makes the perfect continuative [1a]. The

continuative perfect corresponds to the previously mentioned cases 1) and 3) of present

perfect use described by Quirk et al. (1985). The non-continuative perfect, on the other hand,

corresponds to case 2). The continuative perfect is restricted to atelic situations, i.e. situations

where there is no inherent end point (Huddleston and Pullum 2002, 142). Therefore, a

sentence like **He has built a house for a year* is not standard and instead one would use the

present perfect progressive form *He has been building a house for a year*.

    When describing an event that occurred at least once at an unspecified point within

the timespan up to TU but is not continuous and does not result in a permanent change of

state, one would use the experiential perfect, also known as the existential perfect

(Huddleston and Pullum 2002, 143). The event described by the experiential perfect may be

located anywhere in the period leading up to and including TU as is seen in the following

examples:

[2]

    a. I **have** now **written** ten pages.

    b. This is the happiest I **have been** all week

    c. Mary **has** only **been** to a hair salon twice.

In example [2a] the situation described takes place at TU as specified by the adjunct *now*, in

[2b] it spans the whole week before TU and in [2c] it could be very far into the past as the

visits to the hair salon may have happened at any point in Mary's life.

According to Quirk et al. (1985, 193), one of the connotations commonly associated with the present perfect is that the result of the action or situation is still in effect at TU. Huddleston and Pullum (2002, 145) call this type of perfect the resultative perfect, which is shown in the following examples:

[3]

    a. My friend **has recovered** from her illness.

    b. I'**ve been** to the dentist.

    c. I'**ve tried** to find a nice pair of shoes to go with my dress.

Example [3a] is the most unambiguous case of a continuing result which begins with the situation described and continues up to TU. In [3b] the result of the action is much vaguer and more open to interpretation – the implied meaning is that since I have gone to see a dentist, my teeth are now in good condition. Example [3b] could, however, just as well be interpreted as an experiential perfect. Indeed, Huddleston and Pullum (ibid.) emphasise that the experiential and resultative perfect are not mutually exclusive as many verb phrases can have both an experiential and a resultative interpretation. Example [3c] does not imply a continuing result but instead a continued *lack* of result and could also be interpreted as an experiential perfect.

Another connotation often associated with the present perfect is recency (Quirk et al. 1985, 193). In many cases the use of present perfect instead of simple past carries the implication that the event occurred recently:

[4]

    a. I **learned** to whistle.

    b. I'**ve learned** to whistle.

In example [4a] the learning is further removed from TU than in [4b]. This implication of recency may explain why the present perfect is often used when reporting a piece of news

(e.g. *Have you heard? Emily has moved to Paris!*) (ibid.; Huddleston and Pullum 2002, 145). When the time period in which the situation described is located is not explicitly stated, it is usually assumed to be the shortest timespan up to TU that is in agreement with the semantics and pragmatics of the sentence (Declerck 1991, 97). Therefore, it would be inappropriate to respond to the question *Have you had lunch?* with *Yes, every day* as the implied timespan of the question is *today*. The term Huddleston and Pullum (2002, 145-46) use to describe this type of present perfect is *perfect of recent past*.

[5]

      a. I **have found** a new favourite tv series.

      b. He **has** recently **left** the country.

The perfect of recent past often appears with the adjuncts *recently* and *just* (Huddleston and Pullum 2002, 146) as illustrated in example [5b]. This is not, however, a requirement as seen in example [5a] where the implication even without an adjunct of recency is that the discovery of the new series happened not long ago. According to Huddleston and Pullum (ibid.), it could be argued that the perfect of recent past is an unnecessary label as the experiential and resultative categories are broad enough to cover all non-continuative uses of present perfect. Both examples [5a] and [5b] can certainly be interpreted as having an ongoing result – in [5a] the speaker now has a new favourite series and in [5b] the person referred to is no longer in the country. However, Huddleston and Pullum (ibid.) go on to argue that the recency category adds an important new dimension not taken into account by the experiential and resultative categories and is thus equally as valid as the other two.

The perfect aspect can also be combined with the past tense to form the past perfect which is usually used to describe events that were already in the past at some predetermined point before TU (Quirk et al. 1985, 195). Like the present perfect, the past perfect places the event somewhere within a timespan up to a determined point in time. However, unlike the

present perfect in which that point is TU, the past perfect defines it as some point anterior to TU (Ibid., 195-96). The point in question (T_1) is often explicitly stated in the rest of the sentence by an adverbial phrase as shown in the following examples:

[6]

    a.   He had already left **when I got there**.

    b.   Wendy had finished the whole series **before summer vacation ended**.

According to Biber et al. (2021, 467) this accounts for c. 70% of all occurrences of past perfect verb phrases. The location of T_1 can, however, also rely completely on context clues which may occasionally make it difficult to identify the intended time reference (ibid.).

Though the past participle is located in the past time-sphere whereas the present perfect is located in the present time-sphere, the two share most of their other qualities. Similar to the present prefect, the past perfect also holds the three meanings of state, event or habit, though these are not in relation to TU but rather T_1 (Quirk et al. 1985, 196):

[7]

    a.   She **had lived** in Finland her whole life.

    b.   I **had caught** the flu and so could not attend the wedding.

    c.   They **had visited** the library more times that year than ever before.

Example [7a] has a meaning of *state continuing up to T_1* – the timespan lasts from birth to T_1, which in this case is not discernible from the sentence alone. [7b] depicts an event which happened at an undefined time in the span leading up to T_1 – the speaker may have caught the flu at any point leading up to the wedding. Lastly, example [7c] shows a continuous habit leading up to T_1.

Often the distinction between past perfect and past simple is not quite as clear as that between present perfect and past simple. The past perfect and past simple can in some cases

be used more or less interchangeably, which is especially true in clauses introduced by *after* (Quirk et al. 1985, 196):

> [8]
>
> a. I made the bed after my cat **had woken up**.
>
> b. I made the bed after my cat **woke up**.

There is no significant difference in meaning between [8a] and [8b] and therefore both the past perfect and past simple form are acceptable. However, there are also many instances in which one form is preferred over the other. For example, when continuative or resultative meaning needs to be conveyed the past participle is the preferred form (Declerck 1991, 119). Nonetheless, the past participle is rather rare in speech and aside from *had been*, which is common in all registers, is mostly found in fiction (Biber et al. 2021, 466).

The perfect aspect can be used in conjunction with modal verbs, though with a slightly narrower set of potential meanings than is possible when combining modal verbs with simple verb forms (Quirk et al. 1985, 235). According to Quirk et al. (ibid.), some of the possible meanings of perfect modal constructions include:

1) Possibility

> They **might have been** late if they had left even a minute later.
> They **could have won** if they had tried a bit harder.

2) Necessity

> You **must have been** exhausted.

3) Prediction

> We **will have finished** by the time you get here.

4) Obligation

> You **should have finished** cleaning before the guests arrived.

When combined with the perfect aspect, modal constructions expressing possibility and obligation tend to imply nonfulfillment of the action in question (Quirk et al. 1985, 135), as illustrated in the examples above.

Some modal verbs co-occur with the perfect more often than others. According to Biber et al. (2021, 496), the permission/possibility modals *might* and *may* and the obligation/necessity modals *must* and *should* are the most common with the perfect aspect. However, they are more limited in their use in combination with the perfect than they are with simple verb forms – the former two can only express possibility when combined with the perfect while the latter are mostly used to express logical necessity. The modals *can*, *will* and *shall* are rarely used with the perfect, though their past counterparts *would* and *could* do occur rather frequently (ibid.).

2.1.3 The Passive Voice

The term *voice* refers to a grammatical category which makes it possible to view the action of a clause in two different ways by changing the way semantic roles align with syntactic functions while the facts reported remain unchanged (Huddleston and Pullum 2002, 1427; Quirk et al. 1985, 159-60). The active voice places the active party of the clause in subject position and the passive party or the "receiver" of the action in object position whereas the passive voice places the passive party in subject position (Huddleston and Pullum 2002, 1427; Declerck 1999, 200). The passive voice is marked by auxiliary *be* + past participle, though it should be noted that not all constructions that follow this pattern are passive structures. For instance, *they were surprised when we showed up* does not feature a passive verb phrase but instead copula *be* combined with a participial adjective – a term I will further elaborate on in section 2.1.4.

A passive clause does not require an active party or "agent", but one can be added by means of an internalised complement tied to the verb phrase, introduced by the word *by*.

Based on this, passive structures can be placed in one of two categories: long passive which includes the agent and short passive in which it is omitted (e.g. Huddleston and Pullum 2002, 1428; Biber et al. 2021, 927):

[9]

a. i. My car was stolen by my friend.   ii. My car was stolen.

b. i. His proposal was rejected by the committee.   ii. His proposal was rejected.

The long passives [9ai] and [9bi] both have a direct active counterpart: *my friend stole my car* and *the committee rejected his proposal*, respectively. In contrast, [9aii] and [9bii] illustrate that short passives do not have an exact active voice equivalent. [9aii] can be expressed in the active voice as *someone stole my car*, which is equivalent to the long passive *my car was stolen by someone*. While the distinction here may not be significant, the same cannot be said about [9bii]. The active voice equivalent would be *they rejected his proposal*, which is not pragmatically equivalent to *his proposal was rejected*. The latter implies that the proposal was rejected specifically by the person or entity it was put forward to while the former makes no such distinction. Quirk et al. (1985, 165) claim that this is often the case with passive clauses: though they may be possible they do not always carry a meaning identical to the active equivalent.

Biber et al. (2021, 927) list four categories of finite passive constructions:

a. Stative passive which describes the state resulting from an action rather than the action itself:

The handouts **were scattered** all over the floor.

b. Dynamic passive which describes an action rather than the resulting state:

Scissors **were used** to cut the ribbon.

c. Get-passives which use the auxiliary *get* with the past participle instead of *be*:

I'm glad I **didn't get spoiled** by my parents as a child.

d. Long passives which include an agent:

The exam **was supervised** by two teachers.

The stative passive and copula + participial adjective constructions are very similar and clauses like the one given as an example in 1) are often ambiguous as to which form is in question. I will go into further detail on the distinction between the verbal and adjectival use of the past participle in section 2.1.4. The dynamic passive is more easily distinguishable as a verb form. The get-passive is somewhat restricted in its use in that it is limited to the informal register and dynamic verbs (Declerck 1991, 203, Quirk et al. 1985, 161). It is, however, often preferred in informal language when the subject of the passive clause is seen as having some level of responsibility for the action despite not being the one performing it (e.g. *she got elected head of the student council*) (Huddleston and Pullum 2002, 1442). *Get* is also preferred over *be* when the situation can be seen as having a positive or negative impact on the subject (e.g. *Kim got robbed on her way home from work*) while *be* has a more neutral tone (ibid.).

There are some constraints to the use of both the active and passive voice, though the active voice is considerably less restricted. Some verbs are limited to only one voice. Almost all verbs can take the active voice but there is a small group which must always be in the passive (e.g. *be reputed*, *be rumoured*, *be born*) (Quirk et al. 1985, 162; Huddleston and Pullum 2002, 1435). The group of verbs that are restricted to the active voice is considerably larger. Having no direct or indirect object, intransitive and copular verbs cannot use the passive voice (ibid.). Notably, three of the verbs examined in this thesis (*swell*, *come*, *go*) are intransitive and so cannot take the passive. Additionally, phrasal verbs are quite limited in their use of the passive. Many phrasal verbs only accept the passive in the abstract figurative sense of the word – *the matter was carefully looked into* is acceptable but **the mirror was*

looked into is not. The type of object may also restrict the use of the passive voice. Transitive verbs can have either phrasal or clausal objects and the latter place constraints on the use of passive (ibid., 163).

Though the passive voice is somewhat limited in its use, there are various structures that do permit it. For passivisation to be possible, an active clause must include a complement to the verb phrase to act as the subject of the passive clause. This complement may be either phrasal or clausal (Quirk et al. 1985, 163). Huddleston and Pullum (2002, 1431-35) outline four types of complement that allow the change from active to passive. Firstly, the complement may be the direct object of a monotransitive verb. Most transitive verbs taking only one object permit passivisation, which makes this the most common of the types. There are, however, some monotransitive verbs that are restricted, either in general or in specific senses, to the active voice:

[10]

a. i. That **fits** you perfectly!   ii. *You **are fitted** by that perfectly!

b. i. You **failed** me.   ii. * I **was failed** by you.

As illustrated in examples [10aii] and [10bii], the verbs *fit* (in the sense of "suit" or "be the right size") and *fail* (in the sense of "let down") do not permit the passive voice. However, some senses of these verbs do allow the passive. Compare:

[11]

a. My car **was** recently **fitted** with a new alarm.

b. The exam **was failed** by almost the entire class.

In these senses the passive is perfectly acceptable.

Secondly, according to Huddleston and Pullum (2002, 1432-33), passivisation is allowed by indirect and direct objects of some ditransitive verbs. In theory, ditransitive

actives can be converted into two alternative passive forms with either the direct or indirect object as the subject of the newly formed passive clause (ibid.).

[12]

a. i. My aunt **gave** <u>me</u> this card.   ii. <u>I</u> **was given** this card by my aunt

b. i. My aunt **gave** me <u>this card</u>.   ii. [?]<u>This card</u> **was given** me by my aunt.

Example [12aii], which places the indirect object from [12ai] as the subject of the passive clause, is a so-called first passive while in [12bii] the subject position is taken by the direct object, making it a second passive. While the first passive is fairly common, the second is very rare. In AmE it is not acceptable and even in BrE is limited to a small range of acceptable uses (ibid.).

Thirdly, passivisation is possible when a noun phrase is an object not of the verb but of a preposition as seen in the following examples:

[13]

a. i. They **toyed** <u>with</u> the idea.   ii. The idea **was toyed** <u>with</u>.

b. i. The mayor **lied** <u>to</u> the people.   ii. The people **were lied** <u>to</u> by the mayor

c. i. No one **has sat** <u>in</u> this chair.   ii. This chair **hasn't been** sat <u>in</u>.

Huddleston and Pullum (2002, 1433) call this the prepositional passive, noting that it is mostly used in informal language as it leads to preposition stranding which is usually avoided in the formal register. According to Huddleston and Pullum (ibid.) prepositional passives fall into two main categories: those where the preposition is dictated by the verb or verbal idiom (illustrated in [13aii] and [13bii]) and those where the preposition is not fixed (illustrated in [13cii]). The former do not follow any general rule on whether passive in permissible as this varies from word to word. The latter allow passive only if the verb phrase indicates a significant property or change in a property of the subject (ibid., 1446).

Lastly, the active can be changed into passive if the complement is a subordinate clause:

[14]

a. i. The staff **suggested** <u>that we should come back later</u>.　　ii. <u>That we should come back later</u> **was suggested** by the staff.

b. i. They **haven't decided** <u>whether they'll go</u>.　　ii. <u>Whether they'll go</u> **hasn't been decided**.

As shown in the examples above, some of these passive constructions sound more natural than others. [14aii], while correct in theory, would rarely appear in actual use. Instead, one would use a variant in which the subordinate clause is extraposed and a dummy *it* is added (cf. *it was suggested by the staff that we should come back later*) (Huddleston and Pullum 2002, 1434). The two examples show two different types of subordinate clause: declarative ([14a]) and interrogative ([14b]). Huddleston and Pullum (ibid.) state that both of these clause types usually allow passivisation but go on to list some exceptions which do not have a passive counterpart:

We **complained** that there was no hot water.

They **rejoiced** that the war was finally over.

Nobody **cares** what happens to us.

They **are wondering** whether they made the right decision.

Other types of subordinate clauses limit the use of passive to a much greater extent and can only appear in a passive clause as an extraposed subject (ibid.).

Like the perfect aspect, the passive voice can also be used with modal verbs. The modals *can* and *could* frequently appear with the passive to express a kind of "logical possibility" (Biber et al. 2021, 497). When used with the passive these modal verbs often

avoid identification of the agent of the main verb, which gives the reported action or situation a meaning of being logically possible rather than being possible for a named person or entity:

[15]

a. Many interesting items **can be found** at a garage sale.

b. The plan is good overall, but some improvements **could be made**.

The modals *must* and *should* are also commonly used with the passive in academic prose to express obligation without specifying the person who is obligated to act (ibid.):

[16]

a. There are some additional factors that **must be considered**.

b. It **should be noted** that this is only one of the many possible explanations.

The volition/prediction modals *will*, *would* and *shall*, on the other hand, only rarely appear with the passive as they often express personal volition (ibid.).

2.1.4 Participial Adjectives

As mentioned before, participle verb forms can also function as adjectives. This applies to both the present and past participle, though here I focus on adjectives which use the past participle form as they are the only ones relevant to this study. From here on, when using the term *participial adjective,* I am referring specifically to adjectives using the past participle form.

Most participial adjectives have all the main syntactic functions of adjectives – they can be used in the attributive, predicative and postpositive position (Declerck 1991, 451). Adjectives are attributive when they premodify the head of a noun phrase (Quirk et al. 1985, 417; Huddleston and Pullum 2002, 528):

[17]

      a. She was wearing a **knitted dress**.

      b. He came up to me with an **excited expression** on his face.

In the predicative position adjectives function as subject complements or object complements which are tied to a subject or object with a verb (ibid.):

[18]

      a. <u>He</u> *was* completely **smitten** with her. [subject complement]

      b. I *prefer* <u>my eggs</u> **fried**. [object complement]

Postpositive adjectives, on the other hand, act as postmodifiers to nouns or pronouns with no linking verb (Quirk et al. 1985, 418; Huddleston and Pullum 2002, 528-9). They occur mostly with indefinite pronouns and with nouns only under restricted conditions (Declerck 1991, 454; Huddleston and Pullum 2002, 528-529):

[19]

      a. <u>Something</u> **unexpected** might happen and ruin everything.

      b. Did you find <u>anyone</u> **interested**?

A postpositive adjective can usually be regarded as a reduced relative cause (e.g. *did you find anyone [who is] interested*) (Quirk et al. 418). Adjectives, in general, appear rarely in the postpositive position and are much more frequent in the attributive and predicative position (Huddleston and Pullum 2002, 529). Participial adjectives specifically are particularly frequent in the attributive position (Biber et al. 2021, 527).

While most participial adjectives can be used in all three main positions, certain verbs have some restrictions. Take for example the verbs *depart* and *go*. While *departed* is restricted to the attributive position, *gone* can only occur in the predicative position (Quirk et al. 1985, 413):

[20]

    a. the **departed** guests

       *The guests are **departed**.

    b. *the **gone** guests

      The guests are **gone**.

The predicative use of *departed* in [20a] sounds archaic and in modern days one would use the perfect mood instead: *the guests have departed*.

There is also variation among participial adjectives as regards gradability and copular verbs. Some participial adjectives like *surprised* are gradable and can be used predicatively with a range of copular verbs whereas others, such as *alleged*, are ungradable and occur in the predicative position mostly with *be* (Biber et al. 2021, 527).

It can sometimes be difficult to distinguish between passives and participial adjectives in the predicative position but there are certain ways to help with this. The most reliable way is to determine whether the past participle form should be interpreted dynamically or statively, as participial adjectives always have a stative meaning while verbal past participials usually have a dynamic meaning (Huddleston and Pullum 2002, 1437-8). In other words, a passive construction usually emphasises the action and a participial adjective the resulting state of said action. This method is not, however, infallible as there are cases where both a dynamic and stative interpretation are possible. Consider the following examples:

[21]

    a. The tv was **broken** during the break-in.

    b. The tv has been **broken** for a while.

    c. The tv was **broken**.

In [21a], a dynamic interpretation is the most logical and *broken* can be easily recognised as a passive. Similarly, in [21b] it is clear that *broken* has a stative meaning, making it a participial adjective. [21c], however, is much more ambiguous and without any further

context it cannot be said with any certainty whether a dynamic or stative interpretation is more appropriate. In cases like this where there are no explicit indicators, it is impossible to determine if a past participle form is a verb or an adjective (Quirk et al. 1985, 415).

There are also some grammatical indicators which can be used to determine the status of a past participle form. One of these is premodification by the intensifier *very*. If the past participle form has the premodifier *very*, it is unambiguously adjectival as verbs cannot be modified by *very* alone (Quirk et al. 1985, 414-5; Huddleston and Pullum 2002, 1436). Compare:

[22]

      a. She was **very frightened**.

      b. *She was **very frightened** by him.

In [22a] the word class of *frightened* would be ambiguous without the intensifier *very* which makes it adjectival. [22b] sounds unnatural as the complement *by him* implies a passive construction.

Complementation is another feature to look at to determine if a past participle is verbal or adjectival as verbs and adjectives differ to some extent in their complementation. Notably, only verbs can have direct objects and predicative complements, which makes the following examples unambiguously verbal (Huddleston and Pullum 2002, 79):

[23]

      a. I've lost the watch **given** <u>me</u> by my grandfather. [direct object]

      b. He should be **considered** <u>a liability</u>. [predicative complement]

Furthermore, a *by* phrase complement following a past participle form is often indicative of a long passive construction as adjectives tend to prefer other prepositions as in the following example (Declerck 1991, 452):

[24]

        a. She was **scared** <u>by a barking dog.</u>

        b. She is **scared** <u>of dogs</u>.

It should be noted, however, that this is not always the case as, according to Quirk et al. (1985, 415), there is an increasing acceptance of participial adjectives with *by* agent phrases.

    Another way to test if a past participle form following *be* is verbal or adjectival is substituting *be* with another copular verb. Passive constructions can only use the verb *be* or in some cases *get*, whereas adjectives in the predicative position are also found with other copular verbs such as *seem*, *become* and *remain* (Huddleston and Pullum 2002, 1437). Consequently, if *be* cannot be replaced by other verbs, it can usually be concluded that the past participle form in question is a verb as in the following examples (ibid.):

[25]

        a.   My car <u>was</u> **stolen** by a thug.
           * My car <u>appeared</u> **stolen** by a thug.

        b.   It <u>was</u> **shown** to me by a close friend.
           * It <u>remained</u> **shown** to me by a close friend.

In both [25a] and [25b] *was* cannot be replaced by another copular verb making these examples unambiguously verbal. Without the agentive *by* phrase in [19a], *appeared* could be substituted for *was* (cf. *my car appeared stolen*). That does not, however, mean that *my car was stolen* is adjectival. In fact, though ambiguous on its own, this clause would most likely be interpreted as having a dynamic meaning in most cases. This goes to show that the copula substitution test only works one way – though it can be said that a past participle form is verbal if *be* cannot be replaced with another copular verb, being able to substitute other verbs for *be* does not make a past participle adjectival.

Even with all the aforementioned methods to determine if *be + past participle* is a passive construction or copular *be* with a participial adjective, there remain ambiguous cases where these methods are of no use. These are usually short clauses where both a dynamic and stative interpretation are possible, such as *the vase was broken*. In theses cases the only way to determine the status of the past participle form is to examine the wider semantic context it appears in.

2.2 Past Tense Shifting

As regards verb inflectional patterns, there is evidence to suggest that non-standard language tends to prefer simplified paradigms (Cheshire 1994, 126). This is sometimes called paradigm levelling and it can be observed both synchronically and diachronically (Fertig 2016, 429). Past tense shifting can be said to be a form of paradigm levelling as it simplifies the inflectional paradigm of an irregular verb. In this section I go over past tense shifting as a linguistic phenomenon. I begin by presenting one possible theory which might, in part, explain the tendency to generalise past tense forms to the functions of the past participle in non-standard English. I then go on to give an overview of the findings of previous studies on PTS which are later compared to the results of the present study in section 5.

2.2.1 Natural morphology

Natural morphology was developed from the theory of natural phonology which was pioneered by David Stampe and his wife Patricia Jane Donegan to explain variation and change in the field of phonology. The main thesis of natural phonology is that the development of sound patterns of languages, both in individuals and language communities, are governed by natural factors inherent to human vocalisation and perception (see e.g. Stampe 1984; Donegan and Stampe 1979; Donegan and Stampe 2009). The theory was adapted by Willi Mayerthaler into the field of morphology and then further developed by

Wolfgang Ullrich Wurzel to explain morphological variation and change within language-systems. The basic thesis of natural morphology is that certain morphological features are more prevalent in language, more resistant to language change and acquired by children at an earlier stage, which makes them more "natural" (Mayerthaler 1981, 2).

There is plenty of discussion in the field of natural morphology regarding what constitutes "naturalness" in language. Mayerthaler (1987, 25-7) proposes that the definition of "natural" depends on whether one is discussing principles of universal grammar (UG) or specific language systems. According to Mayerthaler (ibid.), in terms of UG, a scale of naturalness can be created based on how difficult certain patterns or forms are for the human brain to process. The easier to process, the more natural the feature. Mayerthaler lists three principles for determining how natural a form is in terms of universal language parameters. The better the form fulfils the principles, the more natural it is. First is the principle of constructional iconicity. According to this principle, a more semantically marked form should be more complex than a semantically unmarked one (Mayerthaler 1987, 48-9). For instance, the plural form *cats* shows constructional iconicity as it adds the ending -s to the less semantically marked singular form *cat*. According to Mayerthaler (ibid.), a form is:

A. Maximally iconic if there is only an additional marker (e.g. the plural marker -*s* in English

B. Less than maximally iconic if there is modulation and an additional marker to the less semantically marked form (e.g. some irregular verb forms such as *kept*)

C. Minimally iconic if there is only modulation (e.g. the vowel change in the plural form *geese*)

D. Non-iconic if there is no change (e.g. the plural form *sheep*)

E. Counter-iconic if the form shows less complexity than the semantically less marked form (e.g. the mythical creature *djinni* whose plural form is *djinn*)

Anderwald (2009, 41) applies this principle to English verb inflectional paradigms and presents the following scale of iconicity:

A. Maximally iconic: regular verbs which form their past simple and past participle forms with the suffix *-ed*

B. Less than maximally iconic: irregular verbs which form their past simple and past participle forms with a combination of modulation and a suffix (e.g. *keep – kept – kept*)

C. Minimally iconic: irregular verbs which form their past simple and past participle forms with modulation only (e.g. *swim – swam – swum*)

D. Non-iconic: irregular verbs which do not show a distinction between the three forms (e.g. *put – put – put*)

E. Counter-iconic: there are no examples of this in English verb paradigms

As illustrated by Anderwald's scale, all types of irregular verbs are not equally natural according to the iconicity principle.

The second principle for determining the naturalness of a form is the principle of uniform encoding or the principle of one form having only one function (Mayerthaler 1987, 49; Mayerthaler 1981, 34-5). In this sense the irregular verb paradigms which distinguish present simple, past simple and past participle from one another with unique forms should be considered more natural than those which do not. This conclusion, however, does not take into account other factors surrounding the use of these forms in speech. As Anderwald (2011a, 265) states, the fact that verb paradigms which only distinguish between the present and past tense but not between past simple and past participle are so common in English shows that this two-part paradigm is fully sufficient. The constructions that utilise past participle forms always include at least one auxiliary, which eliminates the possibility of ambiguity between the past simple and past participle forms in speech even if they are

identical. In contrast, present simple and past simple forms are not distinguished from one another by the use of auxiliaries and therefore it can be argued that distinct present and past tense forms are necessary to avoid ambiguity (Anderwald 2009, 9; 2011, 265). Distinct past simple and past participle forms, on the other hand, are effectively redundant. According to Anderwald (2009, 10-11), this may be a contributing factor to the "levelling" of the morphological distinction between the past simple and past participle forms of some irregular verbs in non-standard English.

The final principle is the principle of transparency. The meaning of a transparent form can be easily deduced from the meanings of its constituent parts (Mayerthaler 1981, 35; Mayerthaler 1987, 49). This is somewhat tied to the principle of uniform encoding in that a form is more transparent if its constituent parts are mono-functional. Of English verb forms the most transparent are those which express tense with an added ending and the least transparent is the form *went* which is not in any way related to the base word *go*.

While Mayerthaler's works focus on universal factors which make features of language natural, Wurzel's works illustrate how languages develop their own independent systems of naturalness which may overrule universal principles of naturalness. According to Wurzel (1987, 59-61; 1984, 77), universal principles of naturalness do not explain all morphological variation and change. There are examples across languages of one morphological form being greatly preferred over another even if the two possible forms are equally natural according to Mayerthaler's universal principles, with the less natural form even being preferred in some cases (ibid.). In light of this, Wurzel (1984, 81-89; 1987, 60-62) proposes that each language has a unique set of parameters which define what is natural. Any inflectional system in a given language has certain exclusively occurring or clearly quantitatively prevailing structural features which can be called its "system-defining properties" (Wurzel 1987, 64). Any morphological feature of a language may be examined in

terms of how well it corresponds to the paradigm set by these system-defining properties, that is how "system-congruous" it is (ibid., 65). Inflectional systems show a trend towards uniformity and system-congruity and thus lean toward the most natural inflectional pattern (ibid. 65-66).

English verb inflectional patterns can be sorted into different categories by the criterion of distinction or non-distinction between their three forms (Anderwald 2011a, 261):

- Type 1: present ≠ past ≠ past participle (three distinct forms, e.g. *ring – rang – rung*)

- Type 2: present ≠ past = past participle (two distinct forms, e.g. all regular verbs)

- Type 3: present = past participle ≠ past (two distinct forms, e.g. *come – came – come*)

- Type 4: present = past ≠ past participle (two distinct forms, e.g. *beat – beat – beaten*)

- Type 5: present = past = past participle (one distinct form, e.g. *hit – hit – hit*)

Type 2 is clearly the dominant pattern as it is followed by all regular verbs and, based on the list of standard English verbs by Quirk et al. (1985, 115-120), also the majority of irregular verbs. Therefore, it can be said that the pattern *present ≠ past = past participle* is a system-defining feature in English verb inflection, making any paradigm that follows this pattern system-congruous and thus more natural than a paradigm that does not. According to Anderwald (2009, 9), this makes it a strong attractor for a range of irregular verbs. Indeed, Geeraert and Newman (2011, 14) note that the group of verbs following the pattern *present ≠ past = past participle* is not only the largest but is currently undergoing the most notable increase in non-standard varieties. In this sense it can be said that non-standard verb systems which level the distinction between past simple and past participle by substituting the past simple form for past participle or vice versa, are more natural than the standard English system (Anderwald 2011a, 14).

According to Dressler (2017, 463), there are two main predictions which apply to diachronic morphological change with regard to naturalness:

i. The more natural a feature is, the more stable and resistant to morphological change it is

ii. If, of the two morphological options A and B, A is more natural, then change from B to A is more likely to occur than the opposite

The tendency to shift toward a more natural language can be seen, for instance, in the development of more iconic, uniform and transparent plural coding in German, Dutch and English (Shannon 1989, 21). According to Shannon (ibid., 22), there has also been a marked shift throughout the history of the English language towards more iconicity and away from verb forms which utilise only modulation. This does not, of course, mean that language change always strives toward more naturalness or that all common non-standard forms must be the result of the standard form being less natural. As Dressler (2017, 463) states, these are merely predictions which have been found to be relatively accurate. Predicting language change is not, however, a precise science and according to Mayerthaler (1987, 36) "one can predict areas of change, and one is even able to predict the probable chronology and directionality of change of special form classes, but one cannot predict individual instances of change". Furthermore, even though it can be said that, as a general rule, language change happens in the direction of more naturalness, in this regard phonological and morphological naturalness are often in conflict with one another. As phonological naturalness aims at optimal articulation and perception whereas morphological naturalness strives toward optimal grammatical marking, the two often have conflicting interests (Wurzel 1984, 30). Due to this it can be difficult to determine the key factors behind variation and change in morphological features as they may well be extra-morphologically motivated

2.2.2 Previous Studies on Past Tense Shifting

Though past tense shifting is by no means a recent phenomenon, it has received relatively little attention in the field of linguistics. Shifting in the opposite direction, namely the use of

past perfect forms as past simple forms has been more widely researched (see e.g. Anderwald 2009, 2011a, 2011b, 2012; Bybee 1982, 1995; Szmrecsanyi 2012). Anderwald has studied this phenomenon extensively when it comes to Class 5 irregular verbs which form their past simple and past participle forms through vowel change alone. A rather large subset of Class 5 verbs is what Anderwald calls "Bybee verbs" whose past simple and past participle form follow the phonological pattern *consonant/consonant cluster + [ʌ] + velar/nasal* (e.g. *strung, stuck*) (see Bybee 1995, 431). According to Anderwald (2009, 98-101; 2011a, 262-4; 2011b, 106), this pattern is a strong attractor for irregular verbs belonging to the same class but with three-part paradigms where only the past participle form follows the pattern of Bybee verbs (e.g. *drunk, rung*).

Anderwald (2009, 110) uses the Freiburg English Dialect corpus (FRED) to study the use of these past participle forms as past simple forms and finds that, in the case of the five most frequent verbs, the non-standard past simple form (*sunk, drunk, rung, sung, begun*) is used in around or above 40 percent of all cases with the exception of *begin* at only 20 percent. This indicates that, at least in traditional British dialects, this form of past participle shifting is common. Anderwald (2011a, 259-261) argues that this is likely due to a combination of the phonological pattern of the larger group of Bybee verbs attracting the smaller group of verbs with a three-part paradigm and the verbs moving towards the more natural (in the Wurzel sense) pattern of *present ≠ past = past participle*. Though most of Anderwald's work centres around non-standard past simple forms, she does note that it may be interesting to study the opposite phenomenon, namely PTS. However, she considers PTS to appear, at least as regards the group of Class 5 verbs in her studies, on a much smaller scale and to be "largely random" (Anderwald 2012, 159).

There are, however, studies which indicate that PTS may be more common than Anderwald believes, though most are rather small in scale. One such study by Avis (1953)

examines the field records of the Linguistic Atlas of the United States and Canada (LAUSC) to find instances of the past simple form *drank* used as a past participle in different regions and social classes. Avis (ibid., 107) notes that, while grammarians and dictionary writers are in consensus that the correct past participle form is *drunk*, the actual usage of these forms in American speech tells a different story. The study shows that the non-standard form is almost as common as the standard form in the North Atlantic and North Central states and not limited to speakers of lower educational backgrounds. Even in New England, the home of the "prestige dialect", 38.1% of the informants with higher educational backgrounds use the non-standard form with 47.6% preferring the standard form and the rest using both forms (ibid., 108). The non-standard form is especially common in West Virginia where it is used by 83.3% of all informants in the field records (ibid., 109).

The number of informants in this study is, however, quite low compared to most modern corpus studies and so the results should be seen more as directional rather than absolute. Avis's study is also rather old and thus does not represent the most recent trends in language use. It does, however, indicate that, at least with respect to the verb *drink*, PTS has been relatively common in some regions of The United States of America for decades. In light of Avis's study it also seems that in some American dialects the non-standard past participle *drank* may even be as common as the non-standard past simple form *drunk* in traditional British dialects. This is rather interesting as Anderwald (2011a, 254) mentions briefly that the use of past simple forms as past participles in FRED is very rare with only one instance of the past participle *drank*. The possibility of a regional difference in the direction of shifting could be a topic of interest for further study.

The difference of PTS prevalence between American and British English is examined in a study by Van Ostade (2015) which compares the use of the non-standard past participle form *went* in the British National Corpus (BNC) and the Corpus of Contemporary American

English (COCA). The non-standard form does, indeed, seem to be more common in American English with 104 tokens in COCA compared to only 6 tokens in the BNC (Van Ostade 2015, 306-307). It must be noted, however, that even in COCA use of the non-standard *went* is marginal compared to the standard *gone* with 17,591 tokens (ibid.). This is somewhat surprising as in a small-scale questionnaire by Van Ostade to American English speakers of different ages and backgrounds, two thirds of the respondents report having heard the form *have went* before, some not even seeing it as incorrect (ibid., 301-2). The use of *went* in place of *gone* has also been prevalent enough to be included as a common "usage problem" in 10 different American usage guides since 1770 (ibid., 296). The corpus data, however, implies that this phenomenon is much less common than the questionnaire, usage guides or Avis's (1953) study suggest.

Van Ostade (2015) also examines the different morphosyntactic environments in which nonstandard *have went* and standard *have gone* appear in COCA. According to Van Ostade (2015, 307), the standard form is more varied in what forms it can be preceded by. *Have gone* is predominantly preceded by noun phrases (40%), modal verbs (37%) and subject pronouns (10%) while the majority of cases of *have went* are preceded by a modal verb (84%) and only few by other types of elements (ibid.). This suggests that past simple forms used as past participles may have a particularly strong association with modality. This theory is further supported by Bloomer's (1998) observations on a collection of conversations which feature adult speakers of American English in the Lower Michigan and Long Island areas. Bloomer (1998, 221) observes that, of all the possible contexts in which past participle forms can appear, nonstandard forms are only used in three: modals with perfect infinitives, present perfect and past perfect. Out of these, modal constructions are the most common by a large margin – of the 100 instances of PTS, 90 are preceded by a modal verb, 6 are present perfects and 4 are past perfects (ibid.). The passive voice does not appear at all, however Bloomer

(ibid., 222) hypothesises that this is most likely related to the rarity of passive constructions in speech in general. There are also no participial adjectives, though this is hardly unexpected as Bloomer (ibid.) notes: "Indeed, phrases such as *a knew entity* or *That has been knew for a long time* seem doubtful."

While most studies on PTS focus on spoken language, there is one corpus-based study by Geeraert and Newman (2011) which includes non-standard past participle forms in internet language. In their study, Geeraert and Newman examine the use of past tense forms as past participles in COCA, the BNC, and the internet as a corpus by using the Google search engine. Geeraert and Newman's study is larger in scope than those presented previously in this section as it examines an expansive range of verbs of different frequencies. As regards verb frequencies, Geeraert and Newman (2011, 22) find that in all three corpora, higher frequency verbs such as *come*, *take* and *go* seem to be much more resistant to PTS than lower frequency verbs such as *bite* and *beat* whose past simple forms are relatively established as past participles. Geeraert and Newman (ibid.) argue that this may at least partially be explained by Bybee's Conserving Effect. According to Bybee (2006, 715), frequently used forms have stronger memory representations in the brain and are thus more likely to be accessed as whole lexical items rather than as inflected forms. This makes them more resistant to change. Geeraert and Newman (2011, 23) note, however, that while a Conserving Effect certainly seems to exist, it cannot solely explain the distribution of non-standard forms and cannot be used to accurately predict change:

> By itself, the Conserving Effect says nothing about the stages which typify
> the historical course of a change like PTS, e.g., whether PTS spreads through
> the verb system in a linear way by, say, increments of 10% per century or
> whether PTS spreads in an S-curve manner with slower early and final stages
> and rapid middle stages

It is, therefore, difficult to predict how or if PTS might spread in the future, even taking into account the Conserving effect.

Another observation by Geeraert and Newman (2011, 27-28) is that in light of the corpus data, PTS seems to occur particularly often in the presence of modal auxiliaries, as in *we should have went there sooner.* For example, of the 81 instances of *have went* in COCA, 67 (82,7%) appear in modal constructions such as the example above, while only 40,9% of the instances of the standard *have gone* occur with a modal auxiliary (ibid., 27). Geeraert and Newman (ibid.) also list some verbs that, in the web corpus, seem to actually prefer the non-standard form in modal constructions with percent differences compared to the standard past participle form:

> froze (64%), shook (61%), saw (61%), drove (60%), fell (60%), swam (59%), ran (50%), hid (47%), gave (46%), came (43%), went (42%), wrote (40%), chose (39%), took (39%), showed (33%), broke (31%), sang (31%), beat (31%), began (19%), drank (15%)

These results are in line with Van Ostade (2015) and Bloomer (1998) who both argue that PTS most commonly occurs in modal constructions. To conclude, previous research indicates that PTS is more common in low-frequency verbs than high-frequency verbs, appears more in American English than British English varieties and is most of the time accompanied by a modal auxiliary.

3 Data and Methodology

In this chapter I present the methods and data used in this study. First, I give a brief overview of corpus linguistics as a methodological approach and its advantages in the study of linguistic patterns. I then give some information on the iWeb corpus and the reasoning behind the selection of this particular corpus. Finally, I explain in detail the methods used in this study and some issues that needed to be addressed.

3.1 Corpus Linguistics

Corpus linguistics is a methodological approach to the study of language use and variation which involves utilising a large and organised collection of natural text known as a "corpus" (Biber 2010, 159). Studies in corpus linguistics take an empirical approach to the study of language by analysing patterns of use as they appear in natural language and can rely on both quantitative and qualitative techniques in analysing data (Biber et al. 1998, 4). Biber et al. (1998, 5) argue that, while there is a place in corpus linguistics for calculating frequencies of linguistic features, it is "essential to include qualitative, functional interpretations of quantitative patterns".

Corpus linguistics is not, however, a set of universally agreed upon methods for the exploration of language, although it is characterised by some common features (McEnery and Hardy 2012, 1). The most prominent feature of corpus linguistics is the examination of machine-readable texts, usually in such quantities that sorting and analysing the data by hand and eye alone would be impossible in a reasonable timeframe (ibid., 1-2). With the assistance of computational tools these data can be searched through rapidly and without the possibility of human interference (ibid.). McEnery and Hardy (2012, 1) emphasise the importance of corpus linguistics to modern study of language:

> corpus linguistics has the potential to reorient our entire approach to the study of language. It may refine and redefine a range of theories of language. It may also enable us to use theories of language which were at best difficult to explore prior to the development of corpora of suitable size and machines of sufficient power to exploit them. Importantly, the development of corpus linguistics has also spawned, or at least facilitated the exploration of, new theories of language –theories which draw their inspiration from attested language use and the findings drawn from it

Corpora play a crucial role in modern linguistics as they reveal word and form frequencies and patterns of usage across different types of text (McCarthy and O'Keeffe 2022, 126). What corpora cannot reveal, however, are the reasons behind these patterns and frequencies, leaving them to be interpreted by the researcher (ibid.)

Corpus linguistics can be roughly divided into two different approaches: corpus-based and corpus-driven. Corpus-based research uses the natural text and speech data in a corpus to support, refute or refine theories or hypotheses establishes in linguistic literature (McEnery and Hardy 2012, 6). According to Biber (2010, 163), "the goal [of the corpus-based approach] is not to discover new linguistic features but rather to discover the systematic patterns of use that govern the linguistic features recognized by standard linguistic theory". In contrast, corpus-driven research is not based on previous literature but rather relies on corpus data exclusively to formulate theories of language (Biber 2010, 168-169; McEnery and Hardy 2012, 6). As this study uses previous literature on non-standard verb forms and, more specifically, PTS as a basis for its hypotheses, it is under the scope of corpus-based research. However, McEnery and Hardy (2012, 6) question the need for this binary distinction as, in their view, the corpus by itself cannot be used as a basis for linguistic theory, which effectively makes all research under corpus linguistics corpus-based.

There are many benefits to corpus linguistics as a methodological approach in linguistic research. As this approach relies extensively on computers, it allows the use of larger amounts of language data than approaches in which the data must be sorted and analysed manually, resulting in findings with much greater generalisability and validity (Biber 2010, 159-60). Furthermore, the use of computational tools diminishes the likelihood of human error or subjectivity affecting the results. A computer is programmed to make the same analytical choice every time it encounters the same linguistic phenomenon, which ensures high reliability (ibid., 163). Keeping these factors in mind, corpus linguistics is a suitable approach to answer the research questions introduced at the beginning of this thesis. Answering the questions requires collection of a large number of examples of standard and non-standard past participles in natural internet text that must be analysed in terms of

frequencies and concordances, which would be too time-consuming for one person to do manually.

3.2 The iWeb Corpus

The starting point for any corpus study is selecting the best corpus to match the research question(s) (McEnery and Hardy 2012, 6). To study the use of PTS forms on the internet, I needed a web-based corpus with enough data to make both quantitative and qualitative observations on this linguistic feature. Additionally, the corpus needed to be easily searchable with part of speech tags so that the uses of PTS forms can be categorised based on the different grammatical constructions with which they appear. After considering other options I found the iWeb corpus to best suit the needs of this study as it includes a significantly larger number of texts compared to many similar corpora such as the Corpus of Global Web-based English (GloWbE). It is also compiled in a very systematic way and tagged with the CLAWS 7 automatic tagger, making it very user-friendly. Additionally, the iWeb corpus features texts from a wide range of different types of websites unlike some other big web-based corpora such as the News on the Web (NOW) corpus, which specialises in news websites. As the purpose of this study is to examine the use of PTS forms in internet language in general and not tied to a specific register, such specialised corpora were out of question. These are the main reasons I selected the iWeb corpus for this study.

The iWeb corpus consists of approximately 14 billion words from 22,388,141 web pages collected from 94,391 different web sites, making it one of the largest web-based corpora. It was compiled in 2018 by the creators of COCA. The texts were found using the Google and Bing search engines and selected systematically to ensure that all websites were

from English-speaking countries and that they contained mostly text and not video or images. The process of selection is explained in detail on the iWeb website information page[1].

The biggest issue with the iWeb corpus is the lack of information on the texts. While many corpora (e.g. COCA, NOW) give additional information on texts, such as year of publication or domain, the iWeb corpus only lists the website on which the text appears with no further information. As this sort of information is irrelevant to my study, this is not a problem, however it does mean that the iWeb corpus is not suitable for many types of research. If one wanted to examine the use of PTS forms in different web registers, for instance, it would be quite difficult using the iWeb corpus and using the Corpus of Online Registers of English (CORE) or another similar corpus would be more advisable. However, as there is no need in the present study to distinguish between different registers or to make observations on diachronic change in PTS form usage, the information provided in the iWeb corpus is perfectly adequate.

3.3 Data Collection and Analysis

When conducting a corpus study, one must first define what search strings to use in order to obtain the desired results. This includes finding the right balance between *precision* and *recall* (McCarthy and O'Keeffe 2022, 578). A query with maximum recall includes all relevant texts while a query with maximum precision *only* includes relevant hits (ibid.). Consequently, it is impossible to achieve both perfect recall and precision as using search terms general enough to include every relevant hit means that a large amount of the data will be irrelevant. Conversely, if one uses highly specific search criteria, there is a high possibility of missing out on relevant data. The challenge for a researcher, then, is to find a balance between the two.

---

[1] https://www.english-corpora.org/iweb/

To ensure high recall the search string *verb form_vvn* would be the logical choice as it should show all occurrences of the verb form as a past participle. There were, however, some issues with this method because, as Biber et al. (1998, 262) note, there is no automatic tagger that is 100 percent accurate and there are always bound to be some tagging errors in any corpus that uses one of these tagging systems. Some verbs forms showed more accurately tagged results, but with others a rather significant portion of the shown tokens were not past participle forms and, thus, not relevant to this study, as in the following tokens from the search strings *took_vvn* and *went_vvn* respectively:

(3) But then suddenly 2 weeks back his mother called mine and **broke** off the relationship. And he never spoke to me after that... (iWeb, compatible-astrology.com)

(4) She was on TV for the reality show Armed and Famous in 2007 and **went** through an emotional therapeutic attempt to relieve her of her phobia of cats. (iWeb, popcrunch.com)

Both of these examples are tagged as having past participle forms and the results included many others that were similarly mistagged. This meant that a simple search of *verb form_vvn* would result in a large amount of irrelevant data to be sorted through. As the number of tokens for the different verb forms ranged from around 7,000 to over a million, it would have been impossible to manually sort them into relevant and irrelevant tokens. Instead, I performed multiple different queries for each verb to ensure better precision.

As past participle forms usually appear with either the auxiliary *have* (perfect) or *be* (passive and participial adjectives), I needed to search for both the standard and non-standard past participle form of each verb with these auxiliaries in their different forms. This, however, excluded participial adjectives in the attributive position, which needed to be searched separately. The iWeb corpus allows the user to limit searches to a specific part of speech by adding a part of speech tag to the end of a word. By utilising this function, I could search for participial adjectives in the attributive position by adding an adjective tag to the end of each verb and following it with a noun. As some passives are also formed with the

auxiliary *get* instead of *be* I also needed to search for the different forms of *get* with both the PTS and standard form. This gave me eight search strings which should cover most uses of standard and non-standard past participle forms:

- _vh *non-standard form*
- _vh *standard form*
- _vb *non-standard form*
- _vb *standard form*
- GET *non-standard form*
- GET *standard form*
- *non-standard form*_j NOUN
- *standard form*_j NOUN

As the tagging system used by the iWeb corpus also recognises clitics like *'ve, 'd* and *'m* as the auxiliaries *have* and *be* respectively, there was no need to perform separate searches for forms like *I've went*.

However, the queries with the above-mentioned strings excluded some relevant results which had to be searched separately. Reanalysed variants of modal verbs with the clitic *'ve* (e.g. *coulda*, *would of*) are not recognised by the tagging system as containing the verb *have* and consequently were not included in the initial search results. As these variants appear quite frequently on the internet, not including them would have meant leaving out a significant number of relevant tokens. Therefore, I performed a second set of queries with each individual reanalysed variant (every modal verb with *-a* and *of*) with the non-standard and standard past participle forms of each verb. I also decided to search for attributive participial adjectives with a second adjective before the noun with the strings *non-standard form*_j ADJ NOUN and *standard form*_j ADJ NOUN. There can, of course, be even more adjectives separating the participial adjective and the noun but these constructions are quite rare and for the sake of keeping the amount of work within reasonable limits, I decided to exclude them.

After deciding what strings to use to search the corpus, one must determine the methods of sorting and analysing the search results. One of the two central corpus-linguistic tools used in this study is frequency lists. Gries (2009, 12) describes frequency lists as "the most basic corpus-linguistic tool" as they can be used to compare how often different words or word forms appear in the corpus. This was necessary in my study as one of the objectives was to find out how frequent non-standard past participle forms are on the internet compared to their respective standard forms. By adding the number of occurrences of non-standard past participle forms with the auxiliaries *have*, *be* and *get*, as attributive participial adjectives and in reanalysed modal verb constructions, I obtained the total frequency of the non-standard form of each verb. I then similarly added together the occurrences of standard forms to compare the two. For ease of comparison, I added together the number of standard and non-standard forms of each verb to determine the total number of past participle constructions. I then calculated the portion of standard and non-standard forms in percentages. This made comparison between the different verbs more straightforward. It should be noted, however, that this method assumes that the only possible variants of past participle are the standard form and the PTS form, i.e. that there are no other non-standard variants. This is clearly not the case as there could be any number of less frequent non-standard forms. For instance, a search in the iWeb corpus with the string *_vh drunken* gives 188 hits, which include both verbal (5) and adjectival (6) tokens:

(5) Most importantly however, Tim has **drunken** entirely too much gin the night prior and is suffering the consequences. (iWeb, howl.fm)

(6) He stated that the Blackfoot had **drunken** quarrels amongst themselves, so that in a short time they were separated into small parties, afraid to meet. (iWeb, aadnc-aandc.gc.ca)

This illustrates that *drank* is by no means the only possible non-standard past participle form of the verb *drink*. The exploration of all possible non-standard forms was, however, beyond

the scope of this study. I therefore proceeded with the assumption that any other non-standard variants are marginal and not relevant in the framework of this thesis.

After calculating the total frequencies of each standard and non-standard form, I took a closer look at the grammatical constructions they appear in. As explained in section 2.1, the past participle form of a verb is used in three constructions: the perfect mood, the passive voice and participial adjectives. These can be further divided into subcategories – notably, the perfect mood can be combined with different tenses to form the present perfect and past perfect and passive constructions can be divided into *be*-passives and *get*-passives. Furthermore, perfect and passive constructions can occur with modal verbs. In this study, seven grammatical constructions were of interest: the present perfect, the past perfect, perfect modal constructions, *be*-passive, *get*-passive, passive modal constructions and participial adjectives. For each verb, I determined the distribution of these constructions among all occurrences of the non-standard and standard forms respectively by calculating percentages.

Most of these constructions were easy enough to recognise from the frequency view alone. The search string *_vh verb form*, for instance, shows the number of occurrences of the verb with each distinct form of *have*. It is then a simple matter to count the occurrences with past tense forms of *have* to find the number of past participle occurrences. Perfect modal constructions were similarly found with ease as the search string *_vm _vh verb form* showed all results with a modal verb, *have* and the verb form in question, though the number of constructions with reanalysed *modal verb + have* obviously had to be added to the total. Passive modal constructions were found with the same method but with *have* replaced by *be*. The number of tokens featuring the present perfect was calculated by subtracting the total number obtained from the search *_vm_vh verb form* from the number of constructions with present tense forms of *have*.

As the distinction between passive constructions and participial adjectives in predicative position is not always transparent, these two were the most difficult categories to distinguish. For this I decided to use part of speech tags and searched for *_vb verb form_v* to find passive constructions and *_vb verb form_j* to find adjectives. Here, too, some issues arose with the automated tagger, as shown in the following example, which was among the results for the query *broke_j NOUN*:

(7) iTunes became the second biggest music retailer in America, the MacBook Air revolutionized laptop computing, and the iPod and iPhone **broke** sales records, while changing the way users consumed content and communicated with each other. (iWeb, businessnewsdaily.com)

Here *broke* is clearly not an adjective but a verb in the past tense, yet it is tagged as an adjective. While the tagging in the iWeb corpus seems to be mostly reliable, there are bound to be some errors. Nevertheless, due to the sheer amount of data it was not feasible to manually examine every result obtained with the search strings *_vb verb form_v* and *_vb verb form_j* to determine which category they belong to and so I had to mostly rely on the automated tagging. Some verbs were clearly tagged more accurately in the corpus than others, however. For example the search string *shown_j* gave no results. This suggests that in ambiguous cases the automatic tagger has always tagged *shown* as a verb rather than an adjective. Consider the following tokens:

(8) Well, if you look really closely in each sequence, the background in each clip is actually static, meaning that there is no actual movement of anything in the **shown** clips aside from the characters performances. (iWeb, disneyexaminer.com)

(9) GIMP will calculate new dimensions in pixels and scale the **shown** image on the screen to these pixel dimensions. (iWeb, gimper.net)

In these tokens and many others like them, I would interpret *shown* as an attributive adjective. As it has been tagged as a verb in the corpus, however, tokens like these cannot be found with the search string *shown_j*. Consequently, I decided to perform searches for *the verb form NOUN* and *a verb form NOUN* to find these participial adjectives that would

otherwise not appear in the results. Even so, it is important to keep in mind that the results regarding participial adjectives may not be as accurate as the other grammatical categories as there is so much room for interpretation.

When it comes to the search string _vb verb form_v, some of the verbs gave so many hits that were inaccurately tagged that the results would have been significantly skewed if the automatic tagger was trusted blindly. To ensure better accuracy I had a closer look at the tokens found with the search string _vb verb form_v and if it seemed that a significant portion of the tokens were mistagged, I determined the number of inaccurately tagged hits. The amount of data was, however, so large that manually sorting through all of it would not have been possible in a reasonable timeframe. Therefore, if a search with _vb verb form_v gave over 200 occurrences I created a random sample of 200 tokens and counted the instances of mistagging in this sample. Based on this number I calculated by extrapolation how many instances of mistagging one might expect to find among all tokens.

One major flaw with the automated tagger that could not be ignored was that it frequently identifies the clitic *'s* as *is* even in cases when it is shortened from *has*. This, of course meant that, if left unchecked, many tokens containing present perfect constructions would mistakenly be counted towards the number of passive constructions. Consider the following tokens found with the search string _vb showed_v:

(10) They've taken people we miss more, but you have to admit she**'s showed** more spirit than most. (iWeb, tvtropes.org)

(11) She is a proper judo story and today she's fit and well, she**'s showed** her full potential which is absolutely brilliant. (iWeb, stv.tv)

In both these examples *she's showed* is misinterpreted by the tagging system as *she is showed* instead of *she has showed*. As this was prevalent enough in the data to significantly distort the results, it was necessary to manually sort all the tokens containing *'s* into passive and perfect constructions. Again, if there were more than 200 hits, I created a sample of 200

45

random tokens and counted how many present perfect constructions appeared in the sample and calculated by extrapolation how many present perfect constructions one would expect to find among all tokens with *'s*.

A particularly difficult search string was *_vb went_v* which, in addition to mistagged adjectives, gave many hits which were accurately tagged but not actually past participle forms. In fact, the vast majority of tokens found with the search string fell under this category, for example the following tokens:

(12) upon completion of enrollment of your phone or device, you will **be went** an activation code. (iWeb, bccu.org)

(13) What I did **was went** into the CMD prompt and reset my services. (iWeb, sevenforums.com)

(14) He reasons that Maggey can't be the killer because whoever it **was went** right for the hidden wall safe, which only prosecutors know about. (iWeb, court-records.net)

In token 12 *went* is clearly misspelled *sent*, which was quite common among the tokens as the letters *w* and *s* are close to each other on a standard keyboard. In token 13 the personal pronoun *I* is omitted from between *was* and *went*. In token 14 *was* does not act as an auxiliary to *went* but it is instead a part of the relative clause *whoever it was*. Uses like these made up a majority of the hits for *_vb went_v*.

Some issues arise when using frequencies as the basis for corpus data analysis. Firstly, frequency is not a straight-forward matter which can be measured in an absolute sense – one cannot say, for instance, that over 100 occurrences always means that a phenomenon is frequent (McEnery and Hardy 2012, 48). The frequency of a word or linguistic phenomenon can only truly be measured in relation to others in its category – for example, it can be said that *go* is a high-frequency verb because it is frequent compared to other verbs (ibid.). Secondly, frequencies can sometimes be misleading (Gries 2012, 111). According to Gries (ibid., 111-112), the words *lively, keeper* and *HIV*, for instance, are roughly equally frequent in the BNC but *HIV* only appears in 62 of 100 corpus parts whereas

lively and *keeper* appear in 97. This shows that HIV is much more specialised and not as widely distributed. Gries (2012, 112) argues that in many corpus studies it may be beneficial to include information on the dispersion of the items across different corpus parts in addition to frequency data. The iWeb corpus, however, is not divided into parts like COCA and thus obtaining dispersion information would require visiting each website individually to determine the register of the text. This would be too time-consuming and so dispersion across the different internet registers was not taken into account in this study.

In addition to frequency lists, this thesis also makes use of concordancing, which allows the analysis of words and constructions in context. A concordance displays every occurrence of the searched element, word or string of words in the context of the whole sentence it appears in or a user-specified number of words to its left and right (Gries 2009, 16; McEnery and Hardy 2012, 35; McCarthy and O'Keeffe 2022, 115). This is often called KWIC (keyword in context). As one of the aims of this study was to explore the grammatical and semantic contexts in which PTS appears, the use of concordances was necessary to answer the research questions. After calculating the frequencies of the five grammatical constructions with each form of each verb, I made a random sample of 100 tokens from each construction for each past participle form of each verb. I then used the KWIC view to observe in more detail how the two forms are used in each grammatical context.

4. Results

In this chapter I present the findings of the corpus data analysis. First, I show in a table the frequencies of the standard and PTS form of each verb and their percentage shares of all instances of the past participle. I then compare the verbs with one another, presenting observations related to the different verb classes and verb frequencies. Next, I show the distribution of the different grammatical constructions that the standard and PTS form of each verb appear in, going into greater detail on the differences between the two forms as regards

their usage in the different constructions. I do this by presenting any noteworthy observations made on small samples of the different grammatical constructions. However, as some grammatical constructions with the PTS form have a very low number of tokens in the iWeb corpus, the most relevant observation about these constructions is their infrequent occurrence rather than any usage patterns that can be seen in the few examples found.

4.1 Standard and PTS form Frequency

Verb class	Verb	Past Participles	PTS form	Standard form
class 3	show	786,242	10,993 (1.4%)	775,249 (98.6%)
	swell	52,992	3,218 (6.1%)	49,774 (93.9%)
class 4	break	659,985	39,018 (5.9%)	620,967 (94.1%)
	wake	12,775	916 (7.2%)	11,859 (92.8%)
class 5	drink	149,379	2,611 (1.7%)	146,768 (98.3%)
class 7	go	830,427	10,950 (1.3%)	819,477 (98.7%)
Total		2,491,800	67,706 (2.7%)	2,424,094 (97.3%)

Table 2. The distribution of PTS forms and standard forms in the iWeb corpus

As seen in table 2, PTS forms make up a rather small portion of all past participle forms of the selected verbs in the iWeb corpus. Only 2.7 percent of all past participle occurrences feature the PTS form, however the PTS forms of some of the verbs are clearly more frequent than others.

As Geeraert and Newman (2011) suggest, overall verb frequency seems to have an effect on how often the PTS form occurs. In table 2 it is shown that the two verbs with the highest frequency (*show* and *go*) have the smallest percentage of PTS form occurrences

compared to standard form occurrences (1.4% and 1.3% respectively). Conversely, the verbs with the lowest overall frequency (*wake, swell*) have the highest rate of PTS form occurrence (7.2% and 6.1% respectively). *Break* is the only verb that does not fit this pattern of *lower verb frequency > higher relative PTS frequency* as it has the third highest number of past participle occurrences overall but also the second-highest percentage of PTS forms (5.9%).

It is also possible that PTS form frequency is affected by verb class, though to verify this further research on a wider range of verbs from each class would be necessary. As seen in table 2, the two class 4 verbs have a rather high number of PTS form occurrences relative to the standard variant. This is intriguing as one would expect class 3 verbs to use the PTS form considerably more often than others. Most class 3 verbs – *show* and *swell* included – have two alternative forms in standard English. In addition to the original standard form the PTS form has obtained standard status and is now listed in grammars and dictionaries alongside the variant ending in *–(e)n*. This is not shown in the corpus data, however, as the PTS forms of *show* and *swell* are no more frequent compared to the corresponding standard variants than the PTS forms of the other verbs, all of which are only found in non-standard English.

4.2 Class 3 Verbs

As mentioned earlier, verbs in class 3 are exceptional as most of them are listed in grammars and dictionaries with two alternative past participle forms: the standard form which is formed with a possible change in base vowel and *-(e)n* and the PTS form which takes the regular *-ed* ending. This means that *showed* and *swelled* may also be considered standard, though I refer to them as PTS forms and to *shown* and *swollen* as standard forms.

4.2.1 *Show*

	showed		shown		total	
Present perfect	6,478	58.9%	277,932	35.8%	284,410	
					showed 2.3%	shown 97.7%
Past perfect	1,681	15.3%	14,573	1.9%	16,254	
					showed 10.3%	shown 89.7%
Modal perfect	608	5.5%	4,631	0.6%	5,239	
					showed 11.6%	shown 88.4%
Be-passive	1,835	16.7%	420,897	54.4%	478,497	
					showed 0.4%	shown 99.6%
Get-passive	46	0.4%	1,264	0.2%	1,310	
					showed 3.5%	shown 96.5%
Modal passive	345	3.2%	55,420	7.0%	55,765	
					showed 0.6%	shown 99.4%
Participial adjectives	0	0%	532	0.1%	532	
					showed 0%	shown 100%
total	10,993	100%	775,249	100%	786,242	
					showed 1.4%	shown 98.6%

Table 3. Distribution of grammatical constructions with standard and non-standard past participle forms of *show* in the iWeb corpus.

As illustrated by Table 3, *showed* follows the expected distribution of grammatical constructions for a PTS form. Van Ostade (2015), Bloomer (1998) and Geeraert and Newman (2011) all found PTS forms to have a higher-than-expected co-occurrence with modal verbs, which can be seen in the past participle distribution of *show*. Though only 5.5 percent of the occurrences of *showed* as a past participle form in the iWeb corpus are modal constructions, this can be explained by the overall low co-occurrence of the verb *show* with modal auxiliaries. While modal perfect constructions only account for 5.5 percent of all PTS forms of *show* in the corpus, the standard form appears even less frequently with modal verbs in

proportion to all instances of the standard form (0.6%). In fact, of all tokens with modal

perfect constructions with the verb *show* in the iWeb corpus, 11.6 percent feature the PTS

form even though this non-standard form only accounts for 1.4 percent of all occurrences of

show in the past participle. Thus, it can be said that, relative to *shown*, *showed* appears

frequently in modal perfect constructions. The same cannot be said of modal passive

constructions, however, as they make up 7 percent of all instances of the standard form

compared to only 3.2 percent of the instances of the PTS form. Of all modal passives 99.4

percent feature *shown*.

In the iWeb corpus, *showed* as a past participle form most often appears in the present

perfect (59.2%), displaying a significant difference to *shown* (35.9%). The past perfect also

makes up a larger portion of *showed* usage (15.4%) than it does that of *shown* (1.9%). Passive

constructions, on the other hand, make up the majority of tokens of *shown* (61.5%) while they

are not nearly as common with *showed* (19.9%). This may be related to register as the passive

voice is relatively rare in informal registers where the PTS form is more likely to appear. A

slightly larger portion of *get*-passives use PTS forms compared to *be*-passives. The PTS form

does not appear as an adjective. It is noteworthy that *showed* is more evenly distributed

across the different constructions than *shown*, as previous research suggests that PTS forms

are more constrained in what environments they appear in compared to standard forms.

Shown, therefore, would be expected to have a more even distribution of grammatical

contexts than *showed*. However, in the iWeb corpus *shown* only has significant co-occurrence

with the present perfect and the passive whereas *showed* appears relatively frequently in all

contexts excluding participial adjectives.

One interesting observation about the use of the present perfect is that, in the third

person, there is a noticeable contrast in what kinds of subjects *showed* and *shown* appear

with. Both forms appear more in the third person than in any other, third person forms being

used in 68 of the 100 tokens of *showed* and 84 of the 100 tokens of *shown*. But while *showed* is used primarily with human subjects, *shown* appears mostly with inanimate or abstract subjects. Of the 68 instances of third person *showed*, 42 (61.8%) feature a human subject:

> (15) I know I speak on behalf of the rest of the boys and we can not believe we are   still winning these and the sheer devotion **our fans have showed** online for this is absolutely amazing. (iWeb, j-14.com)

> (16) And whatever you make of the 19th and 20th-century legacy of socialism, an idealistic faith that came badly unglued (but came badly unglued (but may not quite be dead, as **Bernie Sanders has showed** us) (iWeb, www.salon.com)

In contrast, only 23 (27.4%) of the 84 tokens of *shown* in third person have a human subject and the standard form seems to instead co-occur mostly with inanimate and abstract subjects (72.6%). This difference is only observed in the present perfect as in the past perfect both forms of *show* have roughly the same number of occurrences with a human subject (*showed* 68.7% and *shown* 70%). One possible explanation for this is found upon closer examination of the present perfect constructions with *shown*. The subjects appearing the most frequently with the standard form in the present perfect are *research* in the singular and *studies* in the plural:

> (17) However, **research has shown** that the best negotiations are the kind that are a win-win for both parties. (iWeb, grouptravel.org)

> (18) Multiple **studies have shown** that downtime " in both long and short periods " increases productivity. (iWeb, asianefficiency.com)

While these subjects do also occasionally appear with the PTS form, they do so considerably less frequently than with the standard form. It may be that the phrases *research has shown* and *studies have shown* have become so ingrained in the language that they are resistant to change and thus prefer the standard form.

When it comes to the use of modal perfect constructions, there are no striking differences between the two past participle forms, and there is only slight variation in which modal verbs each form most frequently appears with, as shown in table 4.

Modal verb	showed		shown		Total	
would	247	41.5%	2,107	45.6%	2,333	
					showed 10.6%	shown 89.4%
should	117	19.6%	625	13.5%	736	
					showed 15.9%	shown 84.1%
could	88	14.8%	627	13.6%	707	
					showed 12.4%	87.6%
may	55	9.2%	461	10.0%	516	
					showed 10.7%	shown 89.3%
might	38	6.4%	271	5.9%	309	
					showed 12.3%	shown 87.7%
must	43	7.2%	267	5.8%	310	
					showed 13.9%	shown 86.1%
will	4	0.7%	244	5.3%	248	
					showed 1.6%	shown 98.4%
can	2	0.3%	3	0.06%	15	
					showed 13.3%	shown 86.7%
shall	2	0.3%	13	0.3%	5	
					showed 40%	shown 60%

Table 4. Distribution of modal verbs in modal perfect constructions with *showed* and *shown*

The order of frequency of the different modal auxiliaries is almost identical for both forms. *Would* is clearly the most used modal auxiliary with both forms while *shall* and *can* only have a few tokens each. The only deviation is in the order of *might* and *must* and even there it is marginal. The most notable difference between the two variants is the use of *will*. With *showed*, *will* has a frequency comparable to *can* and *shall*, whereas 5.3% of *shown* occurrences are with *will*.

Similarly, the distribution of the different modal verbs in modal passive constructions is rather similar for both forms as seen in table 5. Both *showed* and *shown* appear most frequently with *will* and *can* and not very often with other modals. Modal passive constructions clearly show a more pronounced preference for the standard form when

compared to modal perfect constructions, which may be explained by the relatively high co-occurrence of the standard form with the passive voice in general.

Modal verb	showed		shown		Total	
would	12	3.5%	1,598	2.9%	1,610	
					showed 0.7%	shown 99.3%
should	28	8.2%	3,319	6.0%	3,347	
					showed 0.8%	shown 99.2%
could	12	3.5%	1,698	3.0%	1,710	
					showed 0.7%	99.3%
may	5	1.4%	1,898	3.4%	1,903	
					showed 0.3%	shown 99.7%
might	7	2.0%	438	0.8%	445	
					showed 1.6%	shown 98.4%
must	6	1.7%	3,260	5.9%	3,266	
					showed 0.2%	shown 99.8%
will	205	59.8%	31,966	57.7%	32,171	
					showed 0.6%	shown 99.4%
can	52	15.2%	10,726	19.4%	10,778	
					showed 0.5%	shown 99.5%
shall	16	4.7%	467	0.9%	483	
					showed 3.3%	shown 96.7%

Table 5. Distribution of modal verbs in modal passive constructions with *showed* and *shown*

As regards the passive voice, there is one construction that appears mostly with *showed* and another which appears mostly with *shown*. As a ditransitive verb, *show* can take both a direct and indirect object and either of these can be placed in the subject position in the passive. In the samples of passive constructions with *showed* and *shown*, passives with the indirect object in subject position are more common with *showed*. In fact, only 2 of the 100 tokens of *shown* in passive constructions use the indirect object:

(19) **Respondents need to be shown** something that can not be described in words only (e.g. screen shots from long questionnaires) - when interviews last for half an hour or more. (iWeb, audiencedialogue.net)

(20) Afterward **I was shown** how to make two different drinks out of mixes, one green and one brownenergy or something and it was not terrible. (iWeb, digg.com)

In contrast, there are 16 tokens of *showed* with the indirect object. This suggests that, at least when it comes to the verb *show*, there may be a connection between the PTS form and indirect objects. While passive constructions with the indirect object in subject position are more common among tokens of *showed*, the opposite can be said of passive constructions in the perfect mood such as the following token:

> (21) It also contains very high levels of carotenoids that **have been shown** to significantly reduce the risk of developing breast cancer in postmenopausal women. (iWeb, healthsomeness.com)

This construction barely appears in the sample of *showed* tokens with the passive voice but is prevalent in the equivalent sample of *shown*. This coincides with the high co-occurrence of *research* and *studies* with *shown*. It appears that *shown* may more often be used in the sense of *prove* or *demonstrate* whereas *showed* in a more literal sense or in phrases like *show up* or *show off*.

4.2.2 *Swell*

As illustrated by Table 6, *swelled* is most frequently used in the present perfect but also often occurs in the past perfect and as a participial adjective. *Swollen* shows a comparably similar number of occurrences of both the present and past perfect with the present perfect having a slight preference for the standard form and the past perfect for the PTS form. In section 4.1 it was noted that it is rather unexpected that the PTS forms of both Class 3 verbs have such a low frequency compared to the standard variants when the use of either form is acknowledged as correct by grammars and dictionaries. In Table 4 it can be seen, however, that as regards *swell*, the PTS variant is either the preferred form or very close in frequency to the standard form in all but one category. The impression of the comparatively low frequency of *swelled* is caused entirely by the immense number of instances of *swollen* as an adjective and therefore only observing the distribution of the standard and PTS form among all instances of past participle use can be misleading.

	swelled		swollen		Total		
Present perfect	1,475	45.8%	1,760	3.5%	3,235		
					swelled 46%	swollen 54%	
Past perfect	644	20%	560	1.1%	1,204		
					swelled 53.5%	swollen 46.5%	
Modal perfect	80	2.5%	45	0.09%	125		
					swelled 64.3%	swollen 35.7%	
Be-passive	204	6.3%	125	0.3%	329		
					swelled 62%	swollen 38%	
Get-passive	0	0%	0	0%	0		
Modal passive	18	0.6%	0	0%	18		
					swelled 100%	swollen 0%	
Participial adjectives	797	24.8%	47,284	95%	48,081		
					swelled 1.7%	swollen 98.3%	
Total	3,218	100%	49,774	100%	52,992		
					swelled 6.1%	swollen 93.9%	

Table 6. Distribution of grammatical constructions with standard and non-standard past participle forms of show in the iWeb corpus.

Much like *showed*, *swelled* shows a higher-than-expected co-occurrence with modal verbs. Modal perfect constructions only account for 2.5% of *swelled* tokens, which is slightly lower than the co-occurrence with *showed*. As with *show*, however, the number of modal perfect constructions overall is low as the standard form also rarely occurs in modal perfect constructions (0.09%) and does not occur in modal passive constructons. Looking at the number of instances of co-occurrence of modal verbs with *swelled* and *swollen* respectively, it can be noted that modal verbs seem to prefer the PTS form as it accounts for 64.3 percent of all modal perfect constructions and all instances of the modal passive. There are no get-

passives with either form. Though the search strings *GET swelled* and *GET swollen* give some hits, they are all participial adjectives.

The distribution of passives and participial adjectives among the two variants is interesting as each construction has a preferred form. *Swell* occurs in the passive relatively rarely as it is mostly used as an intransitive verb in the sense of 'to become larger or rounder than normal' or 'to increase in amount or number'[2]. It does, however, occasionally appear as a transitive verb in the sense of 'to make something increase in number'[3], making it passivisable in certain contexts. When *swell* does occur in a passive clause, it is usually in its PTS form *swelled* (204 tokens) rather than the standard form *swollen* (125 tokens). An overwhelming majority of participial adjectives, in contrast, use *swollen* (47,284 tokens) in lieu of *swelled* (797 tokens).

The preference of *swelled* in the passive voice may be related to the different meanings of *swell* and how they are used with the two variants. As an intransitive verb, *swell* can be used to mean either concrete physical growth in size or, in a more abstract sense, growth in amount or number. Transitive *swell*, however, can only have the latter meaning. In both the present and past perfect samples, most instances of *swelled* also have the second meaning:

(22) The number of students classified this way **has swelled** 50 percent from a decade ago. (iWeb, colorincolorado.org)

(23) When the App Store launched in July 2008, it offered 552 apps but this **had swelled** to 15,000 apps by January 2009. (iWeb, androidauthority.com)

In these tokens *swelled* is used to describe growth in number, but it is also often used to describe a more abstract growth:

(24) And the outcry **has swelled** in recent months in response to threats from Kaiser management to cut 75 nurse positions. (iWeb, nationalnursesunited.org)

---

[2] *Longman Dictionary of Contemporary English for Advanced Learners*, 6th ed., s.v. "Swell."
[3] *Oxford Advanced Learner's Dictionary*, 9th ed., s.v. "Swell"

(25) --justifying his inclusion of a number of maverick scientists on a special panel of experts he formed to advise him on AIDS, which **had swelled** into a massive epidemic in South Africa. (iWeb, nelsonmandela.org)

In the present perfect, 81 of the 100 tokens of *swelled* have one of these more abstract meanings and 71 of 100 tokens in the past perfect. Conversely, *swollen* is mostly used to describe physical growth:

(26) I feel this medicine is making me feel worse and my left leg **has swollen** even more. (iWeb, medschat.com)

(27) Once the plants **have swollen** after their winters rest amounts of water can be increased. (iWeb, worldofsucculents.com)

64 of 100 tokens in the present perfect and 82 of 100 tokens in the past perfect have this physical meaning. It is possible that *swelled* is the preferred form in the passive because it has a stronger connection to the abstract meaning of *swell*, which is the only possible meaning in the passive.

Quirk et al. (1985, 107) also note a difference in the meanings of adjectival *swelled* and *swollen* – while *swollen* is generally accepted in both a literal and metaphorical sense, *swelled* is mostly used metaphorically (as in *swelled head* to suggest conceit). This is, however, only partially supported by the corpus data as the majority of examples of both forms have a literal meaning, mostly referring to different parts of the human body as in the following tokens:

(28) My joints in my fingers were **swelled** and almost to the point I thought they were going to pop. (iWeb, saveourbones.com)

(29) While you are immobile, it is a good idea to elevate your foot, because this allows the fluid to drain away from your **swollen** ankle. (iWeb, wisegeekhealth.com)

The distinction of literal and abstract meaning between the two variants is only noticeable when the modified noun is *head*. The contrast can be seen in the following tokens:

(30) Another gold medal to add to his lapel and more fodder for his already **swelled head**. (iWeb, michaelfairmansoaps.com)

(31) Symptoms include listlessness; a **swollen head**, neck, joints and snood and there can be sudden death. (iWeb, allotment-garden.org)

With most other nouns, however, *swelled* has a literal meaning. The only other exception to this is *swelled* in the figurative sense of 'swelled with emotion' as in the following tokens:

(32) He looked with a dejected, angry countenance, his great heart was **swelled with rage and disdain**. (iWeb, eserver.org)

(33) And my heart was **swelled with indignation**. (iWeb, claremont.edu)

As there are no examples of *swollen* used with this meaning, it is presumably associated mostly with *swelled*. As the sample used in this thesis was relatively small, however, further study would be necessary to make any conclusive observations.

Modal verb	swelled		swollen		Total	
would	27	36%	11	24%	38	
					swelled 71.1%	swollen 28.9%
should	2	2.7%	3	6.5%	5	
					swelled 40%	swollen 60%
could	7	9.3%	0	0%	7	
					swelled 100%	swollen 0%
may	11	14.7%	9	19.6%	20	
					showed 55%	shown 45%
might	8	10.7%	4	8.7%	12	
					showed 66.7%	shown 33.3%
must	7	9.3%	5	10.9%	12	
					showed 58.3%	shown 41.7%
will	13	17.3%	14	30.4%	27	
					48.1%	51.9%
can	0	0%	0	0%	0	
shall	0	0%	0	0%	0	

Table 7. Distribution of modal verbs in modal perfect constructions with *swelled* and *swollen*.

Modal constructions with the verb *swell* are extremely infrequent in the iWeb corpus. Both the standard and PTS form co-occur with the modal perfect less than 100 times. Due to the low number of examples the observations made on the distribution presented in Table 7 should be regarded with caution as when the numbers are very low, even a difference of one token can significantly affect the percentages. This means that any perceived patterns may not be representative of wider use. The different modal verbs are distributed rather similarly with both variants. The most noteworthy differences are the lack of any instances of *could* with *swollen* and the prevalence of *will*, which is the most frequently occurring modal verb with *swollen*. Most modal verbs seem to either appear with both forms equally or prefer *swelled.* The only exception to this is *should* but as *swelled* and *swollen* only have 2 and 3 tokens respectively, it is difficult to say if this is a general trend or caused by chance. Modal passive constructions are even less frequent with *swell* as there are only 18 tokens in this category for *swelled* and none for *swollen*. Though there were more hits for both forms, most of them were mistagged participial adjectives such as the following examples:

(34) Never left in a car, but your right, its hard to get it out of there, like it **may be swelled**. I'll try a different battery when I get home. (iWeb, nikonites.com)

(35) Along with pain and stiffness, the right hand **might be swollen** or red with small, hard bumps below the skin. (iWeb, wisegeek.com)

Due to the low number of tokens it is difficult to say anything definitive about the use of different modal verbs with the passive, but it does seem that modal passive constructions are more common with the PTS form.

4.3 Class 4 Verbs

The past simple forms of verbs in Class 4 are realized with a change in the base vowel and their past participle form with the suffix *-(e)n* with a possible base vowel change.

4.3.1 *Break*

	broke		broken		Total	
Present perfect	1,059	2.7%	42,801	6.9%	43,860	
					broke 2.4%	broken 97.6%
Past perfect	646	1.6%	17,270	2.8%	17,916	
					broke 3.6%	broken 96.4%
Modal perfect	380	1.0%	4,450	0.7%	4,830	
					broke 7.9%	broken 92.1%
Be-passive	812	2.1%	88,250	14.2%	89,062	
					broke 0.9%	broken 99.1%
Get-passive	431	1.1%	6497	1.1%	6,928	
					broke 6.2%	broken 93.8%
Modal passive	77	0.2%	24,431	3.9%	24,508	
					broke 0.3%	broken 99.7%
Participial adjectives	35,613	91.3%	437,268	70.4%	472,881	
					broke 7.5%	broken 92.5%
Total	39,018	100%	620,967	100%	659,985	
					broke 5.9%	broken 94.1%

Table 8. Distribution of grammatical constructions with standard and non-standard past participle forms of *break* in the iWeb corpus.

Broke and *broken* have rather similar distributions of grammatical constructions, as seen in Table 8. Both forms occur overwhelmingly as participial adjectives and rarely with modal verbs. It can still be said that modal constructions have a high relative frequency with *broke* as they are the most common of all the constructions in comparison to *broken*. Though this is in line with previous findings on the grammatical contexts PTS appears in, as regards the use of PTS forms as participial adjectives, *broke* does not display expected results as the overwhelming majority (91.3%) of *broke* tokens are adjectival. *Broken* has a slightly more even distribution than *broke* as participial adjectives do not account for as large a percentage

of all *broken* tokens. A significant portion of *broken* examples feature the *be*-passive while with *broke* this construction is not as frequent. *Get*-passives, however, have a higher ratio of PTS form occurrences to standard form occurrences (6.2% / 93.8%) compared to *be*-passives (0.9% / 99.1%).

The prevalence of participial adjectives among the instances of *broke* is explained by the additional meaning it has acquired as an adjective, unrelated to the meaning of verbal *broke*. *Broke* in the sense of "having no money"[4] is in frequent use and even has a separate entry in most dictionaries (see e.g. *Longman Dictionary of Contemporary English for Advanced Learners* [LCE], s.v. "Broke[2]"; *Oxford Advanced Learner's Dictionary* [OLD], s.v. "Broke). This "variant of conventional *broken*" is by no means a new phenomenon as it has been reported as early as 1661 but is especially widespread in modern language (*The Concise New Partridge Dictionary of Slang and Unconventional English* [PDS], s.v. "Broke"). In the iWeb corpus *broke* is used as an adjective mostly in the sense of "lacking funds"– 73 of the 100 sample tokens have this meaning. Some of these include the phrases *go broke* with the meaning of "can no longer operate because there is no money" and *go for broke* meaning "to take big risks when you try to achieve something"[5]:

(36) No one believed him but we got the dust bowl and many farmers **went broke** and moved to California. (iWeb, iagenweb.org)

(37) So I decided to **go for broke** and pay the $50+ to get Dr G's book shipped to where I am. (iWeb, rawfoodsupport.com)

What is remarkable about the use of *broke* as an adjective in the iWeb corpus is that it is quite often in the attributive position, as in the following tokens, despite both LCE and OLD stating that it does not appear before a noun:

(38) If you're a man and you're **a broke slob**, you're not getting an even decent-looking girl. (iWeb, thisistrouble.com)

---

[4] *Oxford Advanced Learner's Dictionary*, 9th ed., s.v. "Broke."
[5] *Longman Dictionary of Contemporary English for Advance Learners*, 6th ed., s.v. "Broke[2]."

(39) Remember the Beats? **Those broke writers and poets** who flocked here from around the country, coming for the cheap cafes? (iWeb, sfmoma.org)

Though the predicative use of *broke* is more common in the sample, it is interesting that 23 of the 100 tokens are attributive even though, according to dictionaries, *broke* should not occur in the attributive position.

Another unexpected finding is the frequency of *broke* as an adjective in its traditional sense of "broken". Even though this meaning is rarer than the first, it appears in the sample much more frequently than expected both in attributive and predicative position. Of the 100 tokens, 27 use the same meaning as the verb, for example the following tokens:

(40) My left metal brake line is **broke** but my right caliper should at least work right? (iWeb, pelicanparts.com)

(41) Spent too much time/money on tiny drill bits drilling out the bolt, and the **broke** extraction bit. (iWeb, toolmonger.com)

While some of the tokens use *broke* in the literal sense of *broken*, most include the idiom *if it ain't broke, don't fix it*, which is "used to say that you should not try to improve a system, situation etc that is satisfactory"[6]. Some examples also include this idiom in a slightly altered form, such as the following tokens:

(42) We will always have to re-evaluate. # There are times where **"it ain't broke, and we don't need to fix it"**. (iWeb, mirasee.com)

(43) This is where the game will probably stay for a long time. # Riot Games # **If its not broke, don't fix it**. (iWeb, team-dignitas.net)

There are no instances of this saying with *broken*, which suggests that it is used exclusively with the PTS form.

There are not many significant differences between *broke* and *broken* when it comes to the distribution of the different modal verbs across all modal perfect constructions, as seen in Table 9. Both forms appear most often with *would* and rarely or never with *can* and *shall*. *Must* and *should* do, however, appear with *broke* with unexpectedly high frequency – 12.5

---

[6] *Longman Dictionary of Contemporary English*, 6th ed., s.v. "Broke[2]."

modal verb	broke		broken		total	
would	130	34.9%	1,369	30.8%	1,499	
					broke 8.7%	broken 91.3%
should	34	9.1%	235	5.3%	269	
					broke 12.6%	broken 87.4%
could	52	13.8%	642	14.5%	694	
					broke 7.5%	broken 92.3%
may	66	17.7%	1,140	25.7%	1,206	
					broke 5.5%	broken 94.5%
might	33	8.7%	473	10.7%	506	
					broke 6.5%	broken 93.5%
must	58	15.5%	303	6.8%	361	
					broke 16%	broken 84%
will	1	0.3%	251	5.7%	252	
					broke 0.4%	broken 99.6%
can	0	0%	15	0.3%	15	
					broke 0%	broken 100%
shall	0	0%	8	0.2%	8	
					broke 0%	broken 100%

Table 9. Distribution of modal verbs in modal perfect constructions with *broke* and *broken.*

percent of all *should* occurrences and 16 percent of all *must* occurrences are with *broke.* It is not clear from the corpus data why these auxiliaries, in particular, are so common with the PTS form. In addition to *should* and *must*, *will* also shows a distinction between the two forms. With *broke*, *will* clearly belongs to the same category of barely used modals as *can* and *shall*, but with *broken* it sees almost as much usage as *must* and is more common than *should*. This is a similar pattern as displayed by *show* in section 4.2.1.

Another intriguing observation on the use of past participle forms of *break* in modal perfect constructions is the prevalence of reanalysed variants of *modal verb + 've* with *broke*. Particularly common is the variant in which *'ve* is reanalysed as *of*:

(44) Well come to find out the screw **must of broke** one of the reluctor teeth. (iWeb, svtperformance.com)

There are also examples of another variant where *'ve* is reanalysed as *-a*:

(45) I mean a baseball bat **coulda broke** that brick in the video guys do it with their heads too thatis sharp.. very.. (iWeb, trueswords.com)

Of the 380 modal perfect constructions with *broke*, 42 (11%) include a reanalysed variant of *'ve* whereas this only applies to 24 (0.7%) of the 4450 modal constructions with *broken*. Register may, once again, be an explaining factor as these variants are found mostly in the informal register much like past tense shifting.

Modal verb	broke		broken		Total	
would	12	15.6%	1,368	5.6%	1,380	
					broke 0.9%	broken 99.1%
should	9	11.7%	1,250	5.1%	1,259	
					broke 0.7%	broken 99.3%
could	7	9.1%	1,413	5.8%	1,420	
					broke 0.5%	broken 99.5%
may	0	0%	936	3.8%	936	
					broke 0%	broken 100%
might	0	0%	644	2.6%	644	
					broke 0%	broken 100%
must	1	1.3%	1,564	6.4%	1,565	
					broke 0.06%	broken 99.94%
will	14	18.2%	4,611	18.9%	4,625	
					broke 0.3%	broken 99.7%
can	34	44.1%	12,161	49.8%	12,195	
					broke 0.3%	broken 99.7%
shall	0	0%	484	2.0%	484	
					broke 0%	broken 100%

Table 10. Distribution of modal verbs in modal passive constructions with *broke* and *broken*.

Broke rarely occurs in modal passive constructions whereas *broken* does so quite frequently, as illustrated by table 10. The distributions of the modals are, once again, quite similar, though the PTS form has a slightly higher percentage of co-occurances with *would* and *should* compared to the other modal verbs. For both forms, *will* and *can* are the most common. It is clear that all modal verbs are significantly more common with the standard form as the PTS form accounts for less than 1 percent of occurrences of each modal verb.

4.3.2 *Wake*

	woke		woken		Total	
Present perfect	194	21.2%	3,078	26.0%	3,272	
					woke 5.9%	woken 94.1%
Past perfect	117	12.8%	1,251	10.5%	1,368	
					woke 8.6%	woken 91.4%
Modal perfect	63	6.9%	417	3.5%	480	
					woke 13.1%	woken 86.9%
Be-**passive**	250	27.3%	6,053	51.1%	6,715	
					woke 3.8%	woken 96.2%
Get-**passive**	104	11.3%	625	5.3%	729	
					woke 14.3%	woken 85.7%
Modal passive	3	0.3%	409	3.4%	412	
					woke 0.7%	woken 99.3%
participial adjectives	185	20.2%	26	0.2%	211	
					woke 87.7%	woken 12.3%
Total	916	100%	11,859	100%	12,775	
					woke 7.2%	woken 92.8%

Table 11. Distribution of grammatical constructions with standard and non-standard past participle forms of *wake* in the iWeb corpus.

As shown in Table 11, the distribution of the different grammatical constructions with *woke* is quite even with no one dominating construction. This cannot be said of *woken*, however, as with the standard form *be*-passive constructions clearly prevail over all others (51.1% of all *woken* tokens). *Be*-passives are also the most common construction with the PTS form, though not by as wide a margin as with the standard form, and of all *be*-passives, *woke* only makes up 3.8 percent. *Get*-passives, however, have the highest ratio of PTS form occurrences to standard form occurrences as seen in table 10 (14.3% / 85.7%).

Modal perfect constructions also have a relatively high ratio of PTS form occurrences compared to standard form occurrences (13.1% / 86.9%), as expected. Modal passive constructions, however, have the lowest ratio of PTS form occurrences. The most notable observation as regards table 10 is the unexpectedly high number of instances of *woke* as a participial adjective (185). Participial adjectives account for 22.9 percent of all uses of past participle *woke* and 87.7 percent of all instances of participial adjectives.

In present perfect constructions, the PTS form is almost always used with *up* to form the phrasal verb *woke up*. Of the hundred tokens of present perfect constructions with *woke*, only nine do not feature this phrasal verb. With *woken*, other uses are more common, though still in the minority (20 out of 100), for example the following token:

(46) It seems to accurately work out when I fall asleep, though my walk from bed to shower in the morning isn't enough to convince it that I **have woken** and to stop recording sleep. (iWeb, pebble.com)

This difference between the two forms can also be observed in the past perfect, though it is less pronounced – of the hundred past participle tokens with *woke*, 85 feature the phrasal verb *wake up* while the equivalent number for *woken* is 77.

As regards this feature, the distinction between the PTS and standard form is the clearest in passive constructions. 85 of the 100 *woke* tokens in the passive voice feature *woke up* whereas *woken up* only appears in 69 tokens. This is especially noticeable in long passives. Of all the 36 long passive constructions with *woke* only four feature the verb independently without *up*:

(47) Colleen used to complain about **being woke** *by smell of Beef and Greens cooking at five in the morning back then*. (iWeb, elliotthulse.com)

(48) Also our first night at the hotel we **were woke** *by screaming adolescents early hours of the morning on a school trip overexcited and banging*. (iWeb, fleetwaytravel.com)

(49) I **was woke** one night *by the sound of chickens dying*- - (iWeb, ghostsofamerica.com)

(50) I **was woke** from my dream *by the voices and tread Of a band*. (iWeb, st-patricks-day.com)

With *woken* the distribution is more even – of the 50 long passive tokens 29 feature *woken up* and 21 *woken*. It also seems that long passives are more common with the standard form as they make up half of all passive constructions with *woken*.

Wake is most often used in the literal sense of 'to come out of the state of sleep or unconsciousness; to be roused from sleep, cease to sleep' (intransitive)[7] or 'to rouse from sleep or unconsciousness' (transitive)[8]. It can also be used in the more figurative senses of 'to become conscious or aware of; to become "alive" to' (intransitive)[9] or 'to arouse to the consciousness or enjoyment of' (transitive)[10]. In the perfect mood the figurative meaning appears to be more common with the standard form with 26 tokens compared to 15 tokens with the PTS form in the present perfect, for example the following tokens:

(51) I just hope and pray that events of the last couple of years **have woke** Bears up to what Timmys aims are. (iWeb, followfollow.com)

(52) Apparently, Facebook has **woken up** to the huge potential that streamed video represents and is now getting serious aboutfield. (iWeb, aardvark.co.nz)

In the past perfect this figurative sense is rarer in general as there are no tokens of *woke* with a figurative meaning and only four tokens of *woken* with this meaning.

It is rather interesting that the figurative meaning is more common with the standard form in perfect constructions as the opposite can be said of participial adjectives. *Woken* as a participial adjective is very rare as *awakened* is generally the preferred adjective. This is illustrated by the number of hits for each form as an adjective in the iWeb corpus (110 for *woken* and 4447 for *awakened*). Most of the *woken* occurrences are, in fact, mistagged and the true number of participial adjectives is only 26. Almost all of these use the literal meaning of *wake*, such as the following tokens:

---

[7] *Oxford English Dictionary Online*, s.v. "Wake, v." sense 7a, https://www-oed-com.libproxy.tuni.fi/view/Entry/225173?rskey=AuGtii&result=5&isAdvanced=false#eid
[8] *Oxford English Dictionary Online*, s.v. "Wake, v." sense 8a
[9] *Oxford English Dictionary Online*, s.v. "Wake, v." sense 7d
[10] *Oxford English Dictionary Online*, s.v. "Wake, v." sense 9b

(53) Ride this wooden rollercoaster, and attempt to defeat the ferocious **woken** dragon. (iWeb, eurotunnel.com)

(54) Next it's a **woken** baby because the dogs next door bark all the live long day, diaper changesup, lunch time, play time, pretending to be a super hero time- - (iWeb, thatswhatchesaid.net)

Woke, however, is only used as a participial adjective in the recently developed figurative sense of 'alert to racial or social discrimination and injustice'[11]:

(55) Very light skin, blue eyes, and makes it his business to talk about Black struggles, being **woke**, etc. (iWeb, brandysource.net)

(56) Nothing says " I have sex " like being **woke** enough to acknowledge women in science, technology, engineering and maths. (iWeb, www.gizmodo.com.au)

Among the 100 woke tokens, there are no instances with any other meaning.

Woke and *woken* also differ in what adjective positions they appear in. All instances of *woken* are in the attributive position whereas woke appears in the attributive (token 57) and predicative (token 58) position:

(57) I'm so tired of seeing **woke** twitter complain about the lack of diversity in tv and then letting great shows like the get down flop. (iWeb, pedestrian.tv)

(58) Kudos to you Heineken, you guys are **woke**. Beer can save the world. (iWeb, elevatormag.com)

The distribution of attributive and predicative *woke* is almost even with predicative use being slightly more common.

When it comes to modal perfect constructions, there are some differences between the two past participle forms of *wake* which can be seen in table 11. *Should* and *must* account for a higher percentage of *woke* tokens than *woken* tokens whereas *will* only occurs with *woken*. 28.8 percent of all *should* occurrences and 18.8 percent of all *must* occurrences are with the PTS form. However, as the number of modal perfect occurrences with woke is quite low, these figures should be regarded with caution. On modal passive constructions it is difficult

---

[11] *Oxford English Dictionary Online*, s.v. "Woke, adj.2" sense 2, https://www-oed-com.libproxy.tuni.fi/view/Entry/58068747?rskey=mWZeW4&result=3&isAdvanced=false#eid

to make even tentative observations. As there are only 3 instances of woke in modal passive constructions, no conclusions can be drawn on the differences between the two forms. Suffice it to say that modal passive constructions seem to overwhelmingly use the standard form.

Modal verb	woke		woken		Total	
would	19	30.2%	146	35.0%	165	
					woke 11.5%	woken 88.5%
should	15	23.8%	37	8.9%	52	
					woke 28.8%	woken 71.2%
could	7	11.1%	41	9.8%	48	
					woke 14.6%	woken 85.4%
may	5	7.9%	51	12.2%	56	
					woke 8.9%	woken 91.1%
might	4	6.4%	39	9.4%	43	
					woke 9.3%	woken 90.7%
must	13	20.6%	56	13.4%	69	
					woke 18.8%	woken 81.2%
will	0	0%	47	11.3%	47	
					woke 0%	woken 100%
can	0	0%	0	0%	0	
shall	0	0%	0	0%	0	

Table 12. Distribution of modal verbs in modal perfect constructions with *woke* and *woken.*

4.4 Class 5 Verb: *Drink*

Class 5 verbs mark their past simple and past participle forms with only vowel change. Some verbs in this class have three distinct forms (*drink – drank – drunk*) whereas others use the same form for present simple and past participle forms (*come – came – come*).

Table 13 shows that *drank* is more evenly distributed across the different grammatical constructions than *drunk*. The PTS form most frequently occurs in present perfect constructions and rather often in the past perfect or *be*-passive constructions. Modal perfect constructions are also relatively frequent with *drank* while *get*-passives and participial

adjectives are rare. The *get*-passive rarely occurs with *drink* but when it does, the PTS form seems to be preferred. Though the search string *GET drunk_v* results in 88 hits, these are all mistagged participial adjectives with *get* in the sense of 'get intoxicated'.

	drank		drunk		Total	
Present perfect	1,155	44.2%	2,798	1.9%	3,953	
					drank 29.2%	drunk 70.8%
Past perfect	595	22.8%	1,696	1.2%	2,291	
					drank 26.0%	drunk 74.0%
Modal perfect	221	8.5%	242	0.2%	463	
					drank 47.7%	drunk 52.3%
Be-**passive**	460	17.6%	2,135	1.4%	2,595	
					drank 17.7%	drunk 82.3%
Get-**passive**	22	0.8%	0	0%	22	
					drank 100%	drunk 0%
Modal passive	155	5.9%	883	0.6%	1,038	
					drank 14.9%	drunk 85.1%
Participial adjectives	3	0.1%	139,014	94.7%	139,017	
					drank 0.002%	drunk 99.998%
Total	2,611	100%	146,768	100%	149,379	
					drank 1.7%	drunk 98.3%

Table 13. Distribution of grammatical constructions with standard and non-standard past participle forms of *drink* in the iWeb corpus.

The standard form is predominantly used as a participial adjective – adjectival uses account for 94.7% of all *drunk* occurrences. This is also where the clearest divide between the two forms can be seen as *drunk* is used in 99.998% of all participial adjective tokens. This means that the relatively low frequency of the PTS form compared to the standard form seen

in table 12 (1.7% / 98.3%) can mostly be explained by the predominance of *drunk* as a participial adjective. If only verbal uses of *drink* are taken into account, *drank* accounts for 25.3 percent of all instances of the past participle. The PTS form is especially frequent in modal perfect constructions in which 47.7 percent of all tokens feature *drank*. This is a remarkably large portion for a non-standard variant.

As regards perfect constructions, there are no major differences in meaning between *drank* and *drunk*. One distinction that may be observed is that, in present perfect constructions the continuative habitual meaning of the perfect shown in tokens 59 and 60 seems to be slightly more common with the PTS form:

(59) We '**ve drank** Stumbras 999 for several years and never tire of it's unique taste. (iWeb, theliquorbarn.com)

(60) It kept me from going toxic. I **have drank** it pretty much every morning since. (iWeb, dadamo.com)

Of the 100 *drank* tokens 18 have a continuative meaning while the same can only be said of 6 *drunk* tokens. This could, however, be caused by chance as the sample is rather small. The distinction between the different meanings of the perfect mood are also not clear-cut and therefore these numbers may be debatable.

Among the *drank* tokens there is one that calls for special attention:

(61) If he **had drunk** what I **have drank**, the same would have happened to him as happened to me. (iWeb, scienceandnonduality.com)

In this token, both the standard form and PTS form are used in the same sentence – the standard form in the past perfect and the PTS form in the present perfect. One can only guess what might be behind the selection of one form over the other in each instance of the past participle in this example. On one hand it could simply be a matter of this particular writer associating *drank* with the present perfect and *drunk* with the past perfect. A particularly strong association between *drank* and the present perfect would not be unlikely as 44 percent of all uses of *drank* as a past participle are present perfect constructions, as can be seen in

72

table 13. On the other hand, the choice might be completely unrelated to tense. Though the reasoning behind the use of each form is not clear, it seems unlikely that the choice is entirely random.

Modal perfect constructions occur almost equally with both forms but table 13 illustrates some differences in the distribution of *drank* and *drunk* with the different modal verbs. While most modal verbs have roughly even distributions, *would* and *will* appear to prefer *drunk* while *should* occurs more often with *drank*. The most common modal verb used with *drank* is *must* (27.1%) whereas *will* (1.4%) is clearly the least frequent of the modal verbs that have at least one occurrence.

Modal verb	drank		drunk		Total	
would	41	18.6%	71	29.3%	112	
					drank 36.6%	drunk 63.4%
should	32	14.5%	17	7.0%	49	
					drank 65.3%	drunk 34.7%
could	31	14.0%	24	9.9%	55	
					drank 56.4%	drunk 43.6%
may	37	16.7%	38	15.7%	75	
					drank 49.3%	drunk 50.7%
might	17	7.7%	14	5.8%	31	
					drank 54.8%	45.2%
must	60	27.1%	52	21.5%	112	
					drank 53.6%	46.4%
will	3	1.4%	25	9.9%	28	
					drank 10.7%	drunk 89.3%
can	0	0%	0	0%	0	
shall	0	0%	0	0%	0	

Table 14. Distribution of modal verbs in modal perfect constructions with *drank* and *drunk*.

As shown in table 15, the standard and PTS form have similar distributions when it comes to the different modal verbs used in modal passive constructions. The only notable difference is that drank does not occur with *shall*. That being said, even with the standard form shall is rare (0.8% of all *drunk* tokens). Unlike modal perfects, modal passives predominantly use the standard form.

Modal verb	drank		drunk		Total		
would	3	1.9%	15	1.7%	18		
						drank 16.7%	drunk 83.3%
should	42	27.2%	268	30.4%	310		
						drank 13.5%	drunk 86.5%
could	7	4.5%	29	3.3%	36		
						drank 19.4%	drunk 80.6%
may	4	2.6%	49	5.5%	53		
						drank 7.5%	drunk 92.5%
might	3	1.9%	2	0.2%	5		
						drank 60.0%	drunk 40.0%
must	16	10.3%	56	6.3%	72		
						drank 22.2%	drunk 77.8%
will	6	3.9%	43	4.9%	49		
						drank 12.2%	drunk 87.8%
can	74	47.7%	414	46.9%	488		
						drank 15.2%	drunk 84.8%
shall	0	0%	6	0.8%	6		
						drank 0%	drunk 100%

Table 15. Distribution of modal verbs in modal passive constructions with *drank* and *drunk*.

4.5 Class 7 Verb: *Go*

Go is the only verb in class 7. It is unique among irregular verbs in that its past simple form is completely unrelated to the present simple and past participle form.

	went		gone		Total	
Present perfect	5,190	47.4%	422,583	51.6%	427,773	
					went 1.2%	gone 98.8%
Past perfect	1,741	15.9%	88,994	10.9%	90,735	
					went 1.9%	gone 98.1%
Modal perfect	3,836	35.0%	49,924	6.1%	53,760	
					went 7.1%	gone 92.9%
Be-passive	7	<0.1%	1,709	0.2%	1,716	
					went 0.4%	gone 99.6%
Get-passive	0	0%	8	<0.1%	8	
					went 0%	gone 100%
Modal passive	2	<0.1%	207	<0.1%	209	
					went 0.9%	gone 99.1%
Participial adjectives	174	1.6%	256,052	31.2%	256,226	
					went 0.06%	gone 99.94%
Total	10,950	100%	819,477	100%	830,427	
					went 1.3%	gone 98.7%

Table 16. Distribution of grammatical constructions with standard and non-standard past participle forms of *go* in the iWeb corpus.

As illustrated by table 16, *went* occurs almost exclusively in the three perfect constructions. Passive constructions are extremely rare, although this is mostly due to the low number of passive constructions with *go* in general. In almost all cases, *go* is intransitive and thus cannot appear in the passive. There are only a few phrases such as *go through* or *go over* which can be passivised and so the number of passive occurrences with both forms is low. The *get-*passive is especially rare. Compared to all the other constructions, modal perfects have a higher percentage of PTS form occurrences relative to standard form occurrences, which is in line with previous observations.

The number of occurrences of *went* as a participial adjective is surprisingly high. The adjectival status of *went* in most of these tokens is, however, very much a matter of interpretation. Though I have interpreted these uses of *went* as adjectives as they follow the auxiliary *be* but are not passives, one might argue that they function like perfect constructions with the auxiliary *be* in place of *have*. This *be*-perfect was in common use with verbs of movement up to approximately the 18th century and the adjective *gone* is, in fact, thought to have developed when these perfect constructions were reinterpreted as copular *be* with a predicative adjective (OED Online, s.v. "Gone, adj. and n."). It is, therefore, no wonder that there is some ambiguity between perfect constructions and participial adjectives when it comes to the verb *go*. This ambiguity appears to be prevalent with the PTS form as all of the tokens could well be interpreted as featuring the *be*-perfect:

(62) There was one I saw a while back which I thought **was went** too far. Bad Moms I think it was. (iWeb, moviechat.org)

(63) Since the real F-22 was still in the works, a lot of guesswork **is went** into creating this sim. (iWeb, oldpcgaming.net)

In tokens 62 and 63 *was* and *is* could be replaced with *had* and *has* respectively to create the more familiar *have*-perfect. In the present tense, there may be even more ambiguity as both *has* and *is* can appear as the clitic *'s*, which further blurs the line between participial adjectives in the predicative position and present perfect constructions.

In the sample for *gone*, however, there is only one such ambiguous case, which is a quote from the Bible:

(64) And it came to pass, when they **were gone** over, that Elijah said unto Elisha, Ask what I shall do for thee away from thee. (iWeb, biblewalks.com)

All other tokens use *gone* in the unambiguously stative adjectival meanings of 'that has left or departed; no longer present; consumed, used up'[12] (token 65) or 'departed from life; dead'[13] (token 66):

(65) It is not the greatest image, but you can see the cross hatching is **gone** after one bedding session (iWeb, heeltoeauto.com)

(66) Sorry mum, its still kind of hard to know he s **gone** , and how I never got to meet him (iWeb, ficwad.com)

Went does not seem to appear in either of these senses and instead is used in contexts where it is unclear if the meaning is stative or dynamic.

Modal verb	went		gone		Total	
would	1,444	37.7%	16,340	32.7%	17,784	
					went 8.1%	gone 91.9%
should	842	21.9%	7,005	14.0%	7,847	
					went 10.7%	gone 89.3%
could	984	25.7%	13,556	27.2%	14,540	
					went 6.8%	gone 93.2%
may	212	5.5%	5,015	10.0%	5,227	
					went 4.0%	gone 96.0%
might	159	4.1%	3,981	8.0%	4,140	
					went 3.8%	gone 96.2%
must	191	5.0%	2,438	4.9%	2,629	
					went 7.3%	gone 92.7%
will	4	0.1%	1,496	3.0%	1,500	
					went 0.3%	gone 99.7%
can	0	0%	23	0.04%	23	
					went 0%	gone 100%
shall	0	0%	70	0.1%	70	
					went 0%	gone 100%

Table 17. Distribution of modal verbs in modal perfect constructions with *went* and *gone*.

---

[12] *Oxford English Dictionary Online*, s.v. "Gone, adj. and n." sense 1a, https://www-oed-com.libproxy.tuni.fi/view/Entry/79888?rskey=tZ83Xb&result=2&isAdvanced=false#eid
[13] *Oxford English Dictionary Online*, s.v. "Gone, adj. and n." sense 1b

Table 17 shows the distributions of modal verbs with *went* and *gone* in modal perfect constructions. Both forms most often occur with *would*, *could* and *should* and relatively infrequently with *will*, *can* and *shall*, though *will* is not quite as rare with the standard form as with the PTS form. *Should* and *would* seem to have particularly high co-occurrence with the PTS form as *went* is used in 10.7 percent of all *should* instances and 8.1 percent of all *would* instances. *Should* is especially represented in reanalysed variants of *have went* (109 of 317 tokens). In general, reanalysed variants, especially those which use *of* instead of *have*, account for a higher percentage of *went* modal constructions (8.3%) than *gone* modal constructions (1.0%). Furthermore, when compared to the portion of PTS form occurrences of all modal perfect constructions (7.1%), a significantly higher percentage of all *-a* and *of* variants occur with the PTS form (37.6%). Thus, it can be concluded that reanalysed variants of modal perfect constructions have particularly high co-occurrence with *went*.

5. Discussion

This chapter is dedicated to discussion on the findings of this study. I first summarise the main observations made in the previous chapter and present general trends found in the use of PTS forms in the different grammatical constructions which past participle forms appear in. I then compare my findings to those of previous studies and consider them in the light of the theory of natural morphology.

In section 4.1 it was noted that class 4 verbs have high PTS form frequency relative to the standard form compared to most of the other verbs. Class 3 verbs showed an unexpectedly high preference for the standard form considering that the PTS form is equally recognised as standard. It was, however, observed that at least when it comes to *swell*, this preference is only seen in participial adjectives, which predominantly take the standard form. When considering only verbal uses of the past simple form as a past participle, the two forms

were relatively evenly distributed. As this prevalence of standard form participial adjectives is seen in most of the examined verbs, comparing the frequencies of the two variants in verbal uses and adjectival uses may be worth a closer look.

Verb class	Verb	PTS forms in verbal past participles	PTS forms in adjectival past participles
class 3	show	10,993 (1.4%)	0 (0%)
	swell	2,421 (49.3%)	797 (1.7%)
class 4	break	3,405 (1.8%)	35,613 (7.5%)
	wake	731 (5.8%)	185 (87.7%)
class 5	drink	2,608 (25.2%)	3 (0.002%)
class 7	go	10,776 (1.9%)	174 (0.06%)

Table 18. Frequency of the PTS form compared to the standard form in verbal and adjectival past participles

As illustrated by table 18, most PTS forms are more frequent in verbal uses. The contrast is particularly noticeable with *swell* and *drink*. If one only takes into account adjectival past participles, *swelled* and *drank* are much less frequently used than *swollen* and *drunk*. However, when considering only verbal uses, *swelled* is used in approximately half and *drank* in roughly a fourth of all instances. The only verbs whose PTS forms have a higher relative frequency in adjectival uses are the two class 4 verbs *break* and *wake*. A possible explanation for this uncharacteristically high number of adjectival PTS form occurrences is presented later in this section. On the whole, past tense shifting seems to be a phenomenon that mostly affects past participles in verbal use. This is in line with Bloomer (1993) who did not find any participial adjectives in his study and notes that the use of a PTS form as an adjective sounds unnatural.

The high frequency of *drank* as a past participle in verbal uses is particularly noteworthy as Anderwald (2009; 2011a; 2011b) lists *drink* among the verbs whose past participle form is often used as a past simple form. According to Anderwald, *drink* is one of the verbs attracted by the phonological pattern of so-called "Bybee verbs" and thus is seeing a shift toward the paradigm *drink – drunk – drunk*. The present study along with Avis (1953), however, indicates that the paradigm *drink – drank – drank* may be equally common in nonstandard English. It is unclear if this is caused by regional variation as Anderwald has studied British English dialects and Avis's study focuses on American English. This study features many different varieties of English as the data for the iWeb corpus has been gathered from websites from all English-speaking countries. The direction of shifting as regards the past simple and past participle forms of *drink* in different regional varieties would be an interesting topic of further study.

Based on the corpus data, the general trend as regards PTS forms appears to be that they have the highest frequency compared to standard forms in perfect constructions, are less frequent with the passive voice and see only marginal use as participial adjectives. There are, however, some outliers. *Swelled*, for instance, is the preferred form in both *be*-passive and modal passive constructions. A possible explanation for this was touched upon in section 4.2.2: as *swelled* is more frequently used with a transitive meaning in perfect constructions than *swollen*, it is not surprising that passive constructions, which can only be formed with transitive verbs, would prefer *swelled*.

The two class 4 verbs (*break*, *wake*), also show an unexpectedly high number of instances of the PTS form as an adjective. The vast majority of past participle *broke* occurrences (91.3%) are participial adjectives while adjectival use also accounts for a relatively high percentage of past participle *woke* instances (20.2%). Furthermore, the PTS form is used in 87.7 percent of all occurrences of participial adjective *wake*. This higher-than-

average frequency of adjectival use may be explained by the divergence of the meanings of the two variants. Both *broke* and *woke* as adjectives have, in a sense, become independent from the verbs they are derived from. What is interesting is that there is also a third class 4 verb whose PTS form is often colloquially used as an adjective with a specific meaning. Like *broke* and *woke*, *shook* also has its own entry in OED Online with the definition 'emotionally or physically disturbed, discomposed, upset'[14]. This suggests that there might be something about class 4 verb PTS forms in particular that makes them more likely to spread into wider colloquial use as participial adjectives.

As regards the frequent co-occurrence of modal verbs with PTS forms, the findings of this study align with those of previous studies (Geeraert and Newman 2012; Van Ostade 2015; Bloomer 1998). With all six verbs, modal perfect constructions have a high ratio of PTS form occurrences to standard form occurrences when compared to other grammatical constructions. The same cannot be said of modal passives, however. With all verbs but *swell*, modals passives are one of the constructions with the lowest share of PTS form use. This cannot be explained by the overall low number of passive constructions with PTS forms as if it were, modal passives should still have a higher rate of PTS form occurrence compared to the standard form than *be*-passives without a modal verb. As seen in table 19, however, this is not the case. While with all verbs the share of PTS forms in modal perfect constructions is significantly larger than that of present perfect constructions without a modal verb, a similar trend is not seen in passive constructions. *Swell* excluded, all verbs have either lower or only slightly higher PTS form frequency in modal passive constructions than in non-modal *be*-passive constructions. Therefore, rather than PTS forms occurring particularly frequently

---

[14] *Oxford English Dictionary Online*, s.v. "shook, adj.", https://www-oed-com.libproxy.tuni.fi/view/Entry/178494?rskey=EIoMio&result=5&isAdvanced=false#eid

with modal verbs, it would be more accurate to say that they often occur in modal perfect

constructions.

Verb class	Verb	PTS form % of perfect occurrences		PTS form % of passive occurrences	
		present perfect	modal perfect	*be-*passive	modal passive
class 3	show	2.3%	11.6%	0.4%	0.6%
	swell	46%	64.3%	62%	100%
class 4	break	2.4%	7.9%	0.9%	0.3%
	wake	5.9%	13.1%	3.8%	0.7%
class 5	drink	29.2%	47.7%	17.7%	14.9%
class 7	go	1.2%	7.1%	0.4%	0.9%

Table 19. Percentages of PTS forms in present perfect, modal perfect, *be*-passive and modal passive constructions

It should also be noted that, while the present study found PTS forms to occur

frequently in modal perfect constructions, the standard form is still preferred with all verbs

but *swell*. This is inconsistent with Geeraert and Newman's (2012) study, which found *broke*,

drank and *went* to be the preferred forms with modal auxiliaries. As Geeraert and Newman

used a different method of compiling their web corpus than the one used by iWeb, it is not

unfeasible that they may have ended up with a corpus where past tense shifting is more

widely represented. As Geeraert and Newman do not fo into great detail on how the corpus

was compiled, this cannot be confirmed.

As regards passive constructions, *swell* is an outlier in several ways. Firstly, the

number of modal passives is significantly lower than for any other verb (18) and this may, in

part, affect the results. As the search string *_vm _vb swollen_v* returned 517 hits, a sample

was created to calculate an estimate of how many modal passives there are in total. As the

sample contained exclusively adjectival uses of *swollen*, it was extrapolated that there are no instances of the standard form in a modal passive construction. However, it is possible that some instances were missed, which may significantly affect the percentages when the numbers are this low. Secondly, *swelled* is more frequent in *be*-passives than any other PTS form and is, in fact, preferred over the standard form. Taking these factors into account, the other verbs are more representative of past tense shifting in modal passive constructions.

In modal perfects, *should* appears to have especially high co-occurrence with PTS forms while *will* is very infrequent compared to standard forms. Something else noteworthy about the use of PTS forms in modal perfect constructions is that a large portion of tokens with reanalysed variants of *modal verb + 've* feature PTS forms. When compared to the percentage of PTS forms in all modal perfect constructions, their share in reanalysed variants is much higher with all verbs. With all verbs but *go* the majority of *-a* and *of* variants occur with the PTS form. *Go* has a significantly higher number of reanalysed variant occurrences than the other verbs (842) and *went* is used in 317 (37.6%) of these. This may be due to the fact that these variants are more likely to appear in informal registers where past tense shifting is also presumably the most prevalent.

Another trend possibly explained by the frequency of grammatical constructions in different registers is the higher frequency of PTS forms in *get*-passives in contrast to *be*-passives. *Get*-passives are predominantly found in informal language while the *be*-passive is more common in formal contexts (Huddleston and Pullum 2002, 1442). The corpus data indicates that *get*-passives are generally more likely to feature PTS forms than *be*-passives as shown in table 20. *Swelled* is once again an outlier as it has no occurrences with the *get*-passive but accounts for 62 percent of all *be*-passive occurrences with *swell*. *Went* also does not appear to follow this trend, though it may be due to the vanishingly low number of *went*

occurrences in all passive constructions. All other verbs make up a higher percentage of *get*-passives than *be*-passives.

Verb class	Verb	PTS forms in get-passives	PTS forms in be-passives
class 3	show	46 (3.5%)	1,835 (0.4%)
	swell	0 (0%)	204 (62.0%)
class 4	break	431 (6.2%)	812 (0.9%)
	wake	104 (14.3%)	250 (3.8%)
class 5	drink	22 (100%)	460 (17.7%)
class 7	go	0 (0%)	7 (0.4%)

Table 20. Number and share of PTS forms compared to standard forms in *get-* and *be-*passives

Though the clearest trends that were found in this study relate to the grammatical constructions PTS forms appear in, the corpus data indicates that there may also be some semantic factors guiding the use of PTS forms. The corpus data indicates that *shown* may more often be used in the sense of *prove* or *demonstrate* whereas *showed* is used in a more literal sense or in phrases like *show up* or *show off*. Furthermore, one of the more surprising findings of this study was that in perfect constructions, *showed* seems to occur more with human subjects and *shown* with inanimate subjects. This is largely due to phrases such as *studies have shown* or *research has shown* which mostly use the standard form. This would be an interesting point of further study as the samples used in this study were quite small and a more extensive look into this specific phenomenon could either support these findings or call them into question.

Swell also seems to have some semantic preferences when it comes to the two past participle variants. Both in verbal and adjectival uses, *swollen* is more often used in a literal sense and *swelled* more figuratively as in a *swelled* head or *swelled* with emotion. The opposite is true of *wake* in verbal past participles as the standard form is used more frequently in a figurative sense than the PTS form. In adjectival past participles, however, *woke* is used exclusively with a figurative meaning and *woken* mostly in the literal sense. This appears to be a trend as most PTS forms which are used as participial adjectives tend to predominantly occur with a figurative meaning.

It is unclear if natural morphology is in any way related to past tense shifting. While system-dependant natural morphology may partially account for the tendency of irregular verb inflectional paradigms to move toward the regular pattern of *present ≠ past simple = past participle*, it does little to explain why some verbs show a higher level of PTS than others. The same can be said of system-independent natural morphology. According to system-independent natural morphology, language tends to move towards naturalness, i.e. more iconicity and transparency and more uniform encoding. The less natural a form is, the more likely it should be to change toward a more natural form. This is not seen in the corpus data, however, as the naturalness (in the Mayerthaler sense) of a standard past participle form and the equivalent PTS form has little to do with PTS form frequency. For most verbs the standard form is, in fact, the more natural of the two (e.g. *break*, *wake, go*) or the forms are equally natural (*drink*). The only exception is the class 3 verbs *show* and *swell* whose PTS forms are more iconic, more transparent and use the most common encoding for a past participle form (*-ed*). Nevertheless, only *swell* has a higher-than-average share of PTS form occurrences while *show* is one of the verbs that are least likely to occur in the PTS form. A more likely explanation for the difference in PTS form frequency among the chosen verbs is Bybee's theory of the conserving effect. Among the verbs chosen for this study, the ones with

higher general frequency tended to have less PTS form occurrences compared to standard form occurrences and vice versa. This suggests that higher frequency verbs may be more resistant to PTS. Nonetheless, neither the conserving effect nor natural morphology can explain why PTS forms occur more often in certain grammatical constructions and with particular meanings.

6. Conclusion

In this corpus-based study I have taken a closer look at past tense shifting which has been researched relatively little. In the beginning of this thesis, I outlined three research questions which I aimed to answer through analysing the corpus data. The first of these questions was "How frequent are the PTS forms of the selected verbs compared to their respective standard forms?" Based on the corpus data past tense shifting does not appear to be very frequent on the internet but is not an entirely marginal phenomenon either. The verb which has the lowest frequency of PTS form usage compared to the standard form is *go* whose PTS form only makes up 1.3 percent of all past participle instances. *Wake* has the highest PTS form frequency with *woke* accounting for 7.2 percent of all past participles. Both class 4 verbs *broke* and *woke* have a particularly high percentage of PTS form occurrence which may suggest that class 4 verbs are especially likely to use the past simple form as a past participle. As this study only included two class 4 verbs, however, further study would be necessary to confirm this. Another observation made on the frequency of PTS form usage was that the higher overall frequency a verb has the less the PTS form is used in comparison to the standard form. This may be due to the conserving effect explained earlier.

The second research question asked what grammatical contexts PTS forms most often appear in. Of the seven grammatical constructions examined in this study, modal perfect constructions clearly had the highest rate of PTS, which is in line with earlier research.

Perfect constructions in general had a higher ratio of PTS forms to standard past participles than passive constructions or participial adjectives. Though PTS forms were quite infrequent in passive constructions, they appeared to be more frequently used in *get*-passives than *be*-passives. Past tense shifting was found to be a phenomenon which mostly affects past participles in verbal uses as adjectival past participles were rare. Furthermore, when adjectival uses of the past participle were left out and only verbal uses taken into account, the relative frequency of PTS forms compared to standard forms was higher. An exception to this was *broke* and *woke* which both have a unique meaning as an adjective and thus occur more frequently as participial adjectives.

The third research question was "What other relevant observations can be made based on the corpus data?" Most of these observations had to do with the different meanings the PTS form and standard form appeared to be associated with. Some of the verbs appeared to prefer a specific meaning with the PTS form and another with the standard form. These differences in meaning can be found in more detail in chapter 4.

While this study has shed some light on the relatively unknown phenomenon of past tense shifting, it has limitations. Firstly, this study only gives a very general idea of past tense shifting on the internet. As the iWeb corpus does not give any background information on the websites the texts are from, no distinctions could be made as regards PTS form usage across registers. The use of PTS forms in the different web registers using the classification developed by Biber et al. (2015a; 2015b) may be a good topic for further research. Secondly, this study is limited to past participles in finite verb phrases as past participles in non-finite verb phrases could not be easily found with any search string. Non-finite participle phrases do exist, however:

*My sister is always grumpy **when woken up***

As there was no way to reliably search for these in the iWeb corpus, however, they were left outside the study.

The third and possibly most notable limitation of this study has to do with the unreliability of the automatic tagger in the iWeb corpus. The amount of work required was considerably larger than initially thought due to the necessity of sorting through large numbers of tokens to determine what category they belong in. The CLAWS 7 tagger seemed to have the most difficulty in distinguishing past participles used in passive constructions and as participial adjectives even in unambiguous cases. This was the most prevalent with the verb *drink* as a vast majority of the hits returned by the search string _vb drunk_v were adjectives. Another serious flaw with the tagger was the clitic *'s* which was rarely tagged accurately as *has* or *is*. Due to the unreliability of the tagging in the corpus, the frequencies shown in the frequency view could not be trusted and the results had to be examined more carefully in the context view. As the numbers for some verb forms were very high, however, it was not possible to examine each token individually. For the verb forms which returned a high number of hits, frequencies had to be extrapolated from a smaller sample. This means that some of the frequencies shown in the results section may be inaccurate, though they should point in the right direction.

Though this study has its limitations, it raises some questions which may warrant further study. The distribution of PTS forms across the web registers is one topic worth looking into as it was not covered by this study. Another perspective not accounted for is the diachronic development of past tense shifting. A particularly interesting point of further study would be the regional differences in the use of the inflectional paradigms *drink – drunk – drunk* and *drink – drank – drank*. This could be done using the GloWbE as the corpus data gathered from the internet is classified based on region of origin. Furthermore, there may be cause for more extensive examination of class 4 verbs as the ones chosen for this study

showed interesting results when it comes to overall PTS frequency and particularly PTS forms used as participial adjectives. On the whole, past tense shifting is a prominent enough linguistic feature in non-standard English that it certainly warrants more research than has been done thus far.

Bibliography

Primary Source

Davies, Mark. 2018. The iWeb corpus. www.english-corpora.org/iweb.

Secondary Sources

Anderwald, Lieselotte. 2011a. "Are Non-Standard Dialects More 'Natural' than the Standard? A Test Case from English Verb Morphology." *Journal of Linguistics* 4, no. 2 (2011): 251-274. doi: 10.1017/S0022226710000241.

———. *The Morphology of English Dialects: Verb Formation in Non-Standard English*. Cambridge: Cambridge University Press, 2009.

———. "Norm Vs Variation in British English Irregular Verbs: The Case of Past Tense Sang Vs Sung." 2011b. *English Language and Linguistics* 15, no. 1 (2011): 85-112. doi: 10.1017/S1360674310000298.

———. "Variable Past-Tense Forms in Nineteenth-Century American English: Linking Normative Grammars and Language Change." *American speech* 87, no. 3 (2012): 257–293. doi: 10.1215/00031283-1958327.

Avis, Walter S. "The Past Participle Drank: Standard American English?" *American speech* 28, no. 2 (1953): 106–111. doi: 10.2307/453165

Baron, Naomi S., and Naomi S. Baron. *Always on: Language in an Online and Mobile World*. New York: Oxford University Press, 2008. https://ebookcentral.proquest.com/lib/tampere/detail.action?pq-origsite=primo&docID=497603

Bloomer, RK. "You Shoulda Saw Me: On the Syntactic Contexts of Nonstandard Past Participles in Spoken American English." *American speech* 73, no. 2 (1998): 221–224. doi: 10.2307/455743.

Bybee, Joan. "From Usage to Grammar: The Mind's Response to Repetition." *Language* 82, no. 4 (2006): 711–733.

———. *Morphology: A study of the relation between meaning and form*. Amsterdam and Philadelphia: John Benjamins, 1985.

Bybee, Joan L., and Dan I. Slobin. "Rules and Schemas in the Development and Use of the English Past Tense." *Language* 58, no. 2 (1982): 265–89. doi: 10.2307/414099.

Biber, Douglas. "Corpus-based and corpus-driven analyses of language variation and use." In *The Oxford handbook of linguistic analysis*, edited by Bernd Heine & Heiko Narrog, 159–92. Oxford: Oxford University Press, 2010.

Biber, Douglas, Jesse Egbert, and Mark Davies. 2015a. "Developing a Bottom-up, User-Based Method of Web Register Classification." *Journal of the Association for Information Science and Technology* 66, no. 9 (2015): 1817–1831.

———. 2015b. "Exploring the Composition of the Searchable Web: a Corpus-Based Taxonomy of Web Registers." *Corpora* 10, no. 1 (2015): 11–45.

Biber, Douglas, Stig Johansson, Geoffrey N. Leech, Susan Conrad and Edward Finegan. *Grammar of Spoken and Written English*. Amsterdam: John Benjamins, 2021.

Biber, Douglas, Susan Conrad & Randi Reppen. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press, 1998.

Declerck, Renaat. *A Comprehensive Descriptive Grammar of English*. Tokyo: Kaitakusha, 1992.

Denison, David. "Clues to Language Change from Non-standard English." *German life and letters* 61, no. 4 (2008): 533–545. doi: 10.1111/j.1468-0483.2008.00442.x

Deuter, Margaret, Joanna Turnbull, and Jennifer Bradbury. *Oxford Advanced Learner's Dictionary of Current English*. Ninth edition. Oxford: Oxford University Press, 2015.

Donnegan, Patricia Jane and David Stampe. "Hypotheses of Natural Phonology." *Poznan Studies in Contemporary Linguistics* 45, no. 1 (2009): 1-31. http://ling.hawaii.edu/wp-content/uploads/PSiCL45.1.1-312009.pdf

———. "The Study of Natural Phonology." in *Current Approaches to Phonological Theory*, edited by Daniel A. Dinnsen, 126-173. Bloomington: Indiana University Press, 1979. http://ling.hawaii.edu/wp-content/uploads/1979-StudyofNaturalPhonology.pdf

Dressler, Wolfgang U. "Naturalness and Morphological Change." In *The Handbook of Historical Linguistics*, edited by Brian Joseph and Richard Janda, 461–471. Oxford, UK: Blackwell Publishing, 2017.

Fertig, David. "Mechanisms of Paradigm Leveling and the Role of Universal Preferences in Morphophonological Change." *Diachronica* 33, no. 4 (2016): 423–460. doi: 10.1075/dia.33.4.01fer

Fox, Chris, Roz Combley, and Karen Cleveland-Marwick. *Longman Dictionary of Contemporary English*. 6th edition. Harlow, Essex: Pearson Education Limited, 2014.

Geeraert, Kristina, and John Newman. "I Haven't Drank in Weeks: The Use of Past Tense Forms as Past Participles in English Corpora." In *Corpus-Based Studies in*

Language Use, Language Learning, and Language Documentation, edited by John Newman, Harald R. Baayen and Sally Rice, 11–33, 2011. https://web-p-ebscohostcom.libproxy.tuni.fi/ehost/ebookviewer/ebook/ZTAwMHh3d19fNDEw MjQyX19BTg2?sid=ac062225-475e-4bf8-8ac3-83d8312b81f0@redis&vid=0&format=EB&rid=1

Gries, Stefan Th. "Dispersions and Adjusted Frequencies in Corpora." In *Corpus Linguistics Volume IV: Methods and Applications*, edited by Douglas Biber and Randi Reppen, 111-143. Los Angeles: Sage, 2012.

———. Gries, Stefan Th. *Quantitative Corpus Linguistics with R: A Practical Introduction*. London: Taylor and Francis, 2009.

Huddleston, Rodney D., Geoffrey K. Pullum, and Laurie Bauer. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press, 2002.

Lass, Roger. "Proliferation and option-cutting: the strong verb in the fifteenth to eighteenth centuries." In *Towards a Standard English: 1600–1800*, edited by Dieter Stein & Ingrid Tieken-Boon van Ostade, 81–113. Berlin: De Gruyter, 1994.

Mayerthaler, Willi. *Morphologische Natürlichkeit*. Wiesbaden: Athenaion, 1981.

———. "System-independent Morphological Naturalness." in *Leitmotifs in Natural Morphology*, edited by Wolfgang U. Dressler, 25-58. Amsterdam: John Benjamins, 1987.

McCarthy, Michael J. and Anne O'Keeffe. *The Routledge Handbook of Corpus Linguistics*. Second edition. London: Routledge, 2022. https://www-taylorfrancis-com.libproxy.tuni.fi/books/edit/10.4324/9780367076399/routledge-handbook-corpus-linguistics-anne-keeffe-michael-mccarthy

McCulloch, Gretchen. *Because Internet: Understanding the New Rules of Language*. New York: Riverhead Books, 2019.

McEnery, Tony and Andrew Hardie. 2012. *Corpus Linguistics: Methods, Theory and Practice*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511981395

Oxford English Dictionary Online. Oxford: Oxford University Press. https://www-oed-com.libproxy.tuni.fi/

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. *A Comprehensive Grammar of the English Language*. London: Longman, 1985.

Shannon, Thomas. "On Different Types of Naturalness in Morphological Change." *Language Typology and Universals* 42, no. 1 (1989): 20-33. https://www.researchgate.net/publication/309623267_On_Different_Types_of_Naturalness_in_Morphological_Change.

Stampe, David. "On Phonological Representations." in *Phonologica 1984*, edited by Wolfgang U. Dressler, Hans C. Luschützky, Oscar E. Pfeiffer and John R. Rennison, 287-299. Cambridge: Cambridge University Press, 1987. http://ling.hawaii.edu/wp-content/uploads/DS-1984-002-On-Phonological-Rep.pdf

Szmrecsanyi, Benedikt. *Grammatical Variation in British English Dialects: A Study in Corpus-Based Dialectometry*. Cambridge University Press, 2012.

Van Ostade, Ingrid Tieken-Boon, and Viktorija Kostadinova. "Have Went – an American Usage Problem." *English language and linguistics* 19, no. 2 (2015): 293–312. doi: 10.1017/S1360674315000118

Wurzel, Wolfgang U., *Flexions-morphologie und Natürlichkeit*. Berlin: Akademie-verlag, 1989.

———. "System-dependent Morphological Naturalness in inflection." in *Leitmotifs in Natural Morphology*, edited by Wolfgang U. Dressler, 25-58. Amsterdam: John Benjamins, 1987.