Georg-August-Universität Göttingen

Seminar für Englische Philologie

Language Change and (Ir)regularization

Dissertation to obtain a PhD in Philosophy at the Faculty of Arts in the Georg-August-Universität Göttingen

Presented by

Basima Othman Mahmood Al-Hussein

Supervised by:

Principal supervisor: Prof. Dr. Hedde Zeijlstra

Co-supervisor: Dr. Daniele Panizza

Göttingen, April 2018

Abstract

For the last two decades, three main approaches have dealt with the nature of regular versus irregular aspects of language processing in human mind. According to connectionism, all inflected forms are processed in the associative memory (Rumelhart and McClelland 1986; Plunkett and Marchman 1993; Elman 1999). So, both regulars and irregulars are predicted to display frequency effects as a result of storage. However, rule-based approaches predict that all inflected forms are generated by rules and hence they are predicted to exhibit no frequency effects as an outcome of the computation. In both approaches, regularization processes (e.g. using *binged* instead of *brought* in the past tense) and irregularization processes (e.g. *brang* instead of *brought*) are predicted to be at the similar rates. This bidirectional prediction of verbal changes by approaches of rules and storage is in conflict with the prediction by the dual mechanism approach that defends the necessity of two separate mechanisms for language processing: storage and for irregular processing and rules for regular processing. Irregular verbs can be produced correctly if they are memorized and retrieved successfully before the rule-governed route creates forms of regularization. Nevertheless, the irregularization rate is predicted to be rare compared to the regularization rate. Hence, from the dual mechanism perspective, verbal changes mostly occur unidirectionally, towards regularization only. To date, many studies of language processing cannot offer fully results that undoubtedly approve the predictions of either approach.

The current study will attempt to make a contribution to this debate by investigating recent linguistic developments and movements in English verbal system in Contemporary English. I run a corpus study based on data from the multilingual environment of the internet where language change is expected to be faster than in any monolingual environment. In this study, I aim to explore whether verbal developments and changes are towards regularization only favoring the dual mechanism approach or towards both regularization and irregularization favoring single mechanism approaches. The results of the selected data in the current study suggest that on the synchronic level there is a trend towards regularization, while irregularization processes occur rarely. These findings are compatible with the dual mechanism approach, whereas they speak against the hypotheses suggested by the single mechanism approaches. The results of the diachronic analyses of regularization show that the regularization rate is slightly increasing in the time spans (old span: 0.68% versus new span: 0.85%). However, the verbal changes in the direction of regularization are not statistically different in the two spans. Similarly, the results of the diachronic analysis of irregularization indicate that the verbal changes towards irregularization are very infrequent and have the diachronic tendency to be constant over time. This means that the results of the diachronic analyses of (ir)regularization are incompatible with the tenets of single and dual mechanism approaches. From the dual mechanism perspective, verbal changes are predicted to be non-constant and unidirectional towards regularization only. From single mechanism perspectives, verbal changes are predicted to be bidirectional towards both regularization and irregularization.

Abstract (German version)

In den letzten zwei Jahrzehnten haben hauptsächlich drei Herangehensweisen mit der Natur von regulären versus irregulären Aspekten der Sprachverarbeitung im menschlichen Hirn beschäftigt. Im Konnektionismus werden alle flektierten Formen im assoziativen Gedächtnis verarbeitet. (Rumelhart and McClelland 1986; Plunkett and Marchman 1993; Elman 1999). Daher wird vorausgesagt, dass sowohl reguläre als auch irreguläre Formen durch die Abspeicherung Frequenzeffekte zeigen müssen. Andererseits sagen regelbasierte Herangehensweisen voraus, dass alle flektierten Formen durch Regeln generiert werden und deshalb keine Frequenzeffekte als Verarbeitungsergebnis zeigen müssten. Beide Herangehensweisen sagen voraus, dass Regularisierungsprozesse (z.B. *bringed* statt *brought* im Englischen past tense) und Irregulierungsprozesse (z.B.. *brang* statt *brought*) im selben Maße auftreten. Diese bidirektionale Voraussage bezüglich Wortwandel von Herangehensweisen, die auf Regeln und Abspeicherung basieren steht in Konflikt mit der Voraussage des Dual Mechanism Approach, welcher die Notwendigkeit zweier separater Mechanismen für die Sprachverarbeitung verteidigt: Abspeicherung für die irreguläre Verarbeitung und Regeln für die reguläre. Irreguläre Verben können korrekt produziert werden, wenn sie gespeichert und erfolgreich abgerufen werden können bevor die regelbasierte Route reguläre Formen produzieren kann. Auf jeden Fall wird vorausgesagt, dass die Irregularisienusgsrate geringer ist als die Regularisierungsrate. Deshalb sind, von der Dual-Mechanism-Perspektive aus gesehen, Verbänderungen größtenteils unidirektional in Richtung Regularisierung. Bis heute können viele Sprachverarbeitungsstudien keine Resultate anbieten, die unzweifelhaft die Voraussagen einer der beiden Herangehensweisen beweisen.

Diese Studie versucht, einen Beitrag zu dieser Debatte zu leisten, indem sie derzeitige linguistische Entwicklungen und Bewegungen im Verbsystem des modernen Englisch untersucht. Ich unternehme eine Korpusstudie basierend auf Daten aus dem multilingualen Bereich des Internets, wo Sprachwandel erwartbarerweise schneller ablaufen sollte als in einem monolingualen Gebiet. In dieser Studie ziele ich darauf ab zu erforschen ob Verbentwicklungen und -veränderungen nur in Richtung Regularisierung sind, was für Dual Mechanism spräche, oder in Richtung von Regularisierung und Irregularisierung, was für die einfachen Mechanismen spräche. Die Resultate der untersuchten Daten in meiner Studie deuten an, dass es auf der synchronen Ebene einen Trend zur Regularisierung gibt, und Irregularisierungen nur selten auftreten. Diese Ergebnisse sind kompatibel mit dem Dual Mechanism, und sprechen andererseits gegen die Hypothesen von einfachen Mechanismen. Die Resultate der diachronen Untersuchung der Regularisierung zeigen, dass die Regularisierungsrate über längere Zeit leicht zunimmt (frühere Zeitspanne: 0.68% spätere Zeitspanne: 0.85%). Andererseits sind die Mengen an Verbänderungen in Richung Regularisierung in den beiden Zeitspannen nicht statistisch relevant. Gleichsam zeigen die Resultate der diachronen Analyse, dass die Verbänderungen in Richtung Irregularisierung nur sehr selten sind und die diachrone Tendez haben, über längere Zeit konstant zu bleiben. Das bedeutet, dass die Resultate der diachronen Analyse der (Ir-)Regularisierung mit den Grundannahmen der einfachen und dualen Mechanismen inkompatibel sind. Aus Sicht des Dual Mechanism, werden Verbveränderungen als nicht konstant und unidirektional in Richtung Regularisierung vorausgesagt. Aus Sicht der einfachen Mechanimen werden Verbänderungen als bidirektional in Richtung von sowohl Regularisierung als auch Irregularisierung vorausgesagt.

Acknowledgement

During my doctoral journey at Georg-August-Universität/Göttingen/ Germany, I have received different kinds of help from various people around me. With these words, I hope to express my gratitude to the people who have helped, inspired and supported me on this journey. This dissertation could never have been accomplished without the help and dedication of those people.

First and foremost, I would like to show my thankfulness towards my supervisors Prof. Dr. Hedde Zeijlstra and Dr. Daniele Panizza who agreed to work on this topic and with their guidance and support this dissertation have seen the light of day. Hedde, I feel speechless when I want to show my deepest gratitude to you on both personal and professional sides. Thank you for providing me with invaluable comments and suggestions on many drafts of the chapters of my dissertation and for all the prompt answers throughout the development of this work. Thank you for your incredible kindness, many words of advice, persistent support through hard times. Thank you for your patience to put me on the right track at every stop along the way and to stick it out till the very end. Credit for the statistical analysis of my data goes Dr. Daniele Panizza. Dani, thank you for your precious explanations and ideas which helped me to be familiar with statistics. Thank you for keeping me on track and getting me to the finish line of analyzing the data in the environment of R programming. Whatever strengths can be found in this work are the result of Hedde's and Dani's influence. If there are any errors or shortcomings in this work, they are certainly mine.

I owe you a debt of gratitude for the staff of University of Lund and the European Commission who offered me the Erasmus Mundus scholarship (action 2/EMIIY project) for the past three years to pursue my doctoral dream. Also, I would like to thank the members of the international office at Georg-August-Universität/Göttingen/Germany. I especially want to thank (the angel) Esther Lauer and Sandra Ludwig who made my first steps before and after travelling to Germany much easier with their kind care and valuable guidance. Gratitude also goes to the members of Graduiertenschule für Geisteswissenschaften Göttingen (GSGG) for their reinforcement. I am much indebted to Dr. Nele Hoffmann, the managing director at the central office of GSGG, for her wise support and continuous encouragement when I faced some critical situations during my study. At the department of English Philology/Göttingen, I would also like to offer a warm thank you to Hildegard Farke, Margitta Strüber, Jovana Gajic, Iulia Petrariu, John Coates, Seid Tvica, Carina Kauf, Louise Raynaud and Omran Omran for their valuable academic comments and kind care along the way. Additional thanks go to some people from my home university (Salahaddin university /Iraq/Erbil) who helped me in the procedures of extending my leave study. In this respect, my sincere thanks go to Halwest Zandi, Nada Jabar, Sedeek Zandi, Muzda Ali, Ala Faruq, Joan Jalal, Nergiz Jalal, Daham Samku, Mohammad Kader, Babaz Faisal, Alif Layla Adu and Shahla Umer.

Finally, I wish to acknowledge my gratitude towards my family and close friends for sharing my ups and downs. Special thanks go to my parents, Othman Mahmood and Shaha Zainal, for their unconditional love and for lifting me up when I need it the most. Deeply, I want to thank Adnan Abed (my husband), Mohammad (my son) and Lina (my daughter) for their love, inspiration and patience. Additional thanks go to my sisters Safa, Hayat, Amil and Dalal, and to my close friends Halwest Zandi, Alicja Swiderska and Hania Wilska for their persistent support and encouragement in the peaks and valleys of my journey.

To the beautiful soul of

Media Majeed

Forever missed

Table of Contents

List of Figures

List of Tables

Table of Appendices

Abbreviations

Phrase	Abbreviation
Regular Verbs (like walk-walked)	RVs
Irregular Verbs (like bring-brought)	IVs
Regular Forms (regularized forms of irregular verbs (like bring- bringed)	RFs
Irregular Forms (irregularized forms of irregular verbs (like bring-brung)	IFs

1 Chapter One: Introduction

Linguists, psycholinguists and cognitive scientists have always been captivated by the structure of language in the human mind. It is an issue of a longstanding dispute concerning how linguistic information is mentally processed and represented by the human language faculty; whether rules are actually employed in language processing or whether they are merely descriptive tools that have no mental counterparts are used in this processing. This issue has been acting as the trigger for a great number of theoretical and empirical studies in many disciplines including linguistics and psycholinguistics over the past two decades. This has led to a serious re-evaluation of many known fundamentals regarding language processing.

With an interest in the inner mechanisms, generative grammar (starting with the standard theory after Chomsky 1957–1965) theorizes that the human language faculty is consist of a finite list of lexical items and a computational component that combines these lexical items to form an infinite number of complex phrases and sentences by means of combinatorial rules. For instance, in the case of English past tense, a regular verb is generated by a rule that adds a suffix *–ed* to a verb stem e.g., *play-played*. From a rule-based perspective, all linguistic expressions are produced by means of rules. This implies that these expressions are not predicted to be frequency-insensitive. Nevertheless, advocates of connectionism, starting with Rumelhart and McClelland (1986), focus on the belief that all linguistic (and non-linguistic knowledge) are processed and acquired through a single associative mechanism namely storage in an associative memory. They, therefore, base themselves on associative explanations of the human language capacity and hence predict that any linguistic processing should display sensitivity to frequency as a reflection of storage. Followers of the dual mechanism approach, starting with Pinker and Prince (1988), combine the central features of generative grammar and connectionism. They employ rules and also incorporate the associative component for language processing. According to this approach, regular expressions are generated by rules, while irregular ones are stored in the associative memory (see chapter 2 for more details).

The bulk of theoretical and empirical studies surrounding the above stated single-dual mechanism debate has focused on inflectional morphology, and particularly on first language (L1) processing and acquisition of the English past and perfect inflections. The reason for the prominence of the past and perfect forms is that the inflectional processes within these two forms appear to comprise two descriptively distinct structures (regular and irregular). In Pinker 's (1999) understanding, regular verbs (RVs) are generated by adding *-ed* to verb stems e.g., *talk-talked-talked, play-played-played*. The majority of the English verbs in the past and perfect forms are regular. Pinker (1999) claims that the regular past tense inflection of *-ed* applies to 86% of the 1000 most frequent verbs in English. This regular inflection is productively applied and generalized to a number of different conditions like new and unknown verbs (e.g., *email-emailed-emailed, fax-faxed-faxed, blick-blicked-blicked*) (Berko1958; Pinker 1991, 1999). Irregular English past and perfect inflections, however, are applied unpredictably (in idiosyncratic ways) to roughly 180 stems of verbs e.g., *cut-cut-cut, buy-bought-bought, speak-spoke-spoken*. Pinker (1999) and Lieberman et al. (2007) argue that irregular verbs (IVs) commonly tend to be high-frequent and high-frequency verbs tend to be irregular. This is supported by that facts that in English the top 10 frequent verbs (*be, have, do, go, say, can, will, see, take* and *get*) are all irregular (Lieberman et al. 2007), of the top 30 verbs in the past tense, 22 are irregular (KuCera and Francis 1967) and of the top 200 verbs in the same tense, 76 are irregular (MacWhinney 2000). IVs can be generalized to other IVs only under specific frequency and phonological circumstances (Prasada and Pinker 1993; Weyerts and Clahsen 1994). Therefore, the English past and perfect formations appear to be served by two separate systems that act independently from each other. One can, accordingly, investigate whether two different mechanisms are at work (suggesting the dual-mechanism approach), or whether this detected binary distinction can be explained by means of one single mechanism (suggesting either rule-based or connectionist theories).

But, what kind of evidence would bear on the nature of processing and acquiring past and perfect forms by single-dual mechanism approaches? The strongest evidence comes from word frequency effects in language processing. Word frequency effects can be a way for diagnosis of the storage hypothesis in which the presence of these effects may be a reflection of storing and retrieving inflected forms from the associative memory. The absence of these effects, however, may imply the application of rules. One way to diagnose word frequency effects is to investigate regularization that refers to processes of over-applying the regular suffix *-ed* to IVs e.g., *speak-speaked-speaked, cut-cutted-cutted*. The regularization processes have become the focus of nearly all empirical studies of past and perfect acquisitions. Marcus et al. (1992) and Pinker (1995) observe that the rate of verb regularization is 4.2%. Nevertheless, later studies have found somewhat higher rates. For instance, Yang (2002) reports a rate of 10% and Maslen et al. (2004) presents a rate of 7.8%. In general, a relationship between word frequency and regularization has been attested: IVs with high word frequency tend to have lower rates of regularization than IVs with low word frequency (Pinker 1999; Lieberman et al. 2007; Michel et al. 2011 among other).

Another way to diagnose word frequency effects is irregularization. This involves processes in which IVs are over-applied to other IVs e.g., *cling-clang-clung*, *slink-slank-slunk, think-thank-thunk* along the lines of *ring-rang-rung*. Lignos and Yang (2015) argue that irregularization processes are rarely studied systematically. They also claim that the regularization rate should be very low, even lower than the rate attested in the study of Xu and Pinker (1995) that is about 0.2%. Regularization and irregularization are originally observed in Berko 's (1958) so called *wug* test in which children commonly add the regular suffix *-ed* to novel verbs such as *rick* and *spow,* whereas they rarely over-apply irregularization; only 1 out of 86 children irregularize *bing and gling* in the past tense into *bang*, *glang* in reference to *ring-rang.* Single mechanism approaches (either rule-based or storage-based) assume that the same mechanism triggers the production of regularization and irregularization processes. Hence, these approaches predict that IVs are regularized at the same rate that they are irregularized. This bidirectional prediction of verbal changes contrasts with the prediction made by the dual mechanism approach in which verbal changes mostly occur unidirectionally (towards regularization only). In the dual mechanism approach, IVs are learned and produced correctly if they are memorized and retrieved successfully before the rule-governed route creates forms of regularization. Therefore, IVs with high frequency are more resistant to regularization processes than the ones with low frequency. Nevertheless, irregularization processes are predicted to be very rare and hence the irregularization rate must be lower than the regularization one (see chapter 2 for details).

It is common knowledge that most morphological changes decrease morphological markedness. For instance, it has been attested that in English the number of IVs has diminished over time gradually, as IVs with low frequency are regularized more often than IVs with high frequency (Fries 1940; Pinker 1999; Lieberman et al. 2007; Michel et al. 2011). A lot of English irregular verbs are undergoing regularization in the course of history e.g. *chide-chid-chid, gripe-grope-gripen* and *wrothe-writhen-writhed are changed into chide-chided-chided, gripe-griped-griped* and *writhe-writhed-writhed* respectively (Pinker 1999: 69). Yet, some linguists have objected the view that looks at linguistic changes in the direction of regularization only, as changes in the other direction, the direction of irregularization, have been observed as well (Nübling 2000 Peters 2009 and Fertig 2013). This is due to the fact that diachronically several

RVs have become irregular in English (e.g. *cost-cost-cost, sneak-snuck-snuck, hang-hung-hung, dig-dug-dug, light-lit-lit, catch-caught-caught, kneel-knelt-knelt, make-made-made and wear-wore-worn, ring-rang-rung*).

Despite many scientific publications on the single-dual mechanism debate of language processing, verb (ir)regularization processes are still much debated. These scientific publications mainly focus on language processing in L1. So, it might also be interesting to investigate language processing in a multilingual environment primarily because this environment can accelerate language development and language change. The claim that in the multilingual environment the situation of language change is more rapid has been already emphasized by Crystal (2004). He stresses that language change in the multilingual environment (especially in Internet) goes faster than at any previous time in linguistic history. Nowadays, multilingualism is diffused (Aronin and Singleton 2008; Auer and Wei 2007; Cook 1992; Grosjean 1982, 2010). Grosjean (1982) conjectures that roughly half the world's population is bilingual.

One important factor that induce language change is language contact. Bussman (1998: 260) defines language contact as "a situation in which two or more languages coexist within one state and where the speakers use these different languages alternately in specific situations". However, nowadays language contact in the virtual environment of the internet does not have to imply the coexistence of two languages within one state. Many people who are located in geographically distant locales, who are of different linguistic backgrounds and who might never come into real contact, can easily engage in an interaction that can be seen as a way of distant language contact. Thomason (2003) further argues that language contact may result in language change which can be any kind of linguistic changes that would have been less likely to occur outside a particular contact situation. Moreover, it is well-known that linguistic changes naturally occur slowly. However, I assume that language contact in the internet may speed up processes of linguistic changes. These linguistic changes can be motivated by the nature of this intensely multilingual medium in which many people are virtually trying to communicate with each other and fostering linguistic experiences never seen before (Danet and Herring 2007). For instance, certain words and linguistic expressions may disappear; existed words or neologisms are inflected using various inflectional expressions. In this respect, Crystal (2011: 67) claims that (ir)regularization processes are commonly used in the internet particularly with innovated words. Some people, for instance, prefer to regularize new words e.g., *google-googled, email-emailed, inbox-inboxed, upload-uploaded*), while other tend to inflect them irregularly e.g., *tweet –twat* or *twot, vax-vaxen, bix-bixen*. Even sometimes inflection expressions reflect a mixture of both e.g., *matrix-martrixes* or *matrixen*. I suppose that such linguistic changes processes are at a much faster speed than before in the internet space.

Developing our knowledge of the multilingual mind particularly in the internet as an increasingly multilingual domain and comparing it with hypotheses and findings regarding the monolingual mind will possibly take us a number of steps beyond our contemporary understanding of the architecture of language in the human mind. The current study will attempt to make a humble contribution to this immense body of research by investigating language processing in the multilingual environment. To this end, I will run a corpus study based on data from the internet environment to explore whether, in Contemporary English, verbal developments and changes are towards regularization (unidirectional) or towards both regularization and irregularization (bidirectional). I aim to investigate how well single and dual mechanism approaches fit the selected data of this study. For this purpose, the following research questions have been formulated (see chapter 3 for more details):

- Are IVs generally more frequent than RVs in the past and perfect forms in Contemporary English?
- Do regularization processes take place in Contemporary English? If so, are IVs with low frequency regularized more often than IVs with high frequency in the past and perfect forms?
- Do irregularization processes take place in Contemporary English? If so, are IVs with low frequency regularized more often than IVs with high frequency in the past and perfect forms?
- Do regularization processes occur more frequently in the cases where IVs and their corresponding irregular forms (e.g., *learn-learnt/learned*) show no vowel change in Contemporary English?
- Are verbal changes towards regularization taking place constantly over time in Contemporary English?
- Are verbal changes towards irregularization taking place constantly over time in Contemporary English?

This thesis is set up as follows.

Chapter 2 (Theoretical Background) overviews the theoretical backgrounds for the study focusing on a brief explanation of the notion of agreement and its properties, different views about the syntax-morphology interface in the literature, main theories of morphological processing in addition to findings from various related empirical studies in favour or against single-dual mechanism approaches and finally internal and external forces for verbal changes.

Chapter 3 (Methodology) serves as an introduction to data analysis in which the methodology that will be used to explore verb (ir)regularization processes in the selected corpus is illustrated. It presents the research questions of the current study and the main predictions of single-dual mechanism approaches for morphological processing. It will then sketch out the motivation of choosing the internet environment and WebCorp as a Linguistic Corpus for the present study. Finally, data selection and procedures that will be followed in the analysis of this study will be described.

Chapter 4 (Results and Analysis) offers descriptive statistics that illustrate general overviews in the selected data using tables and different types of graphs for comparative and descriptive purposes. Then, statistical models will be conducted to test the significance of the difference in the results.

Chapter 5 (Discussion, Conclusion and Suggestions for Further Research) offers a discussion of the results obtained from the data of this study regarding word frequency effects of the past and perfect forms in (ir)regularition processes and draws a conclusion in an attempt to collect evidence for/against single-dual mechanism approaches of morphological processing. Lastly, suggestions for further research are offered.

2 Chapter Two: Theoretical Framework

In this chapter, I will present the main theoretical backgrounds that I need in this study to investigate whether the verbal changes in English are unidirectional (moving towards regularization) or bidirectional (both moving towards both regularization and irregularization). This chapter consists of three major sections.

Section 2.2 covers a brief explanation of the notion of agreement and its properties. In addition, different views about the syntax-morphology interface in the literature will be discussed. In section 2.3, I review the main arguments of a longstanding debate in linguistics and psycholinguistics that relates to how linguistic information is processed by the human language faculty: *are all linguistic processes taken care of by one single mental mechanism (either a rule-based system or an associative system) or by a dual mental mechanism (a rule-based system and an associative system)*? I begin the review by presenting single-mechanism models that rely on rules only (Chomsky and Halle 1968; Halle and Mohanan 1985; Yang 2006), and then move on to present single-mechanism models of associative memory only (Rumelhart and McClelland 1986; McClelland and Patterson 2002). Finally, I discuss the dual-mechanism models that combine the core features of the two previous models (Pinker and Prince 1988; Marcus et al. 1995; Pinker 1999; Pinker and Ullman 2002). In the third major section, I survey main sources of language change: internal and external factors.

2.1 What is agreement?

Agreement is a significant and prevalent phenomenon in natural human languages. It recognizes and identifies that elements in the sentence are linked or should be interpreted together (Bock et al. 2001). It refers to a variety of different types of relationships that may match the constituents of a particular syntactic construction, like subject–verb or modifier–head configurations. For example, in the present tense in English, the regular verb *arrive* shows agreement with its subject in number and person by receiving the third person singular -*s* of in the following sentence:

1. The train arrive-s
   train **3.SG** arrives- **3.SG**
   The train arrives.

Agreement is defined as "systematic covariance between a semantic or formal property of one element and a formal property of another" (Steele 1978: 610 cited in Corbett 2006:4). Corbett (2006) shows that the controller (like the subject) is the first element that determines agreement; it is typically nominal in nature, while the target is the element that is determined by agreement and it may typically be verbs or adjectives. The property in which the target covaries with the controller is called a feature, like person and number features, which may in turn have certain values (first, second, or third for person, and singular or plural for number), as in the example of figure 1. The syntactic environment in which the agreement occurs is called the domain of agreement (for instance, a phrase or clause). Finally, the factors that have indirect effects on the agreement (such as word order) are called the conditions under which the agreement takes place. All these terms of agreement are depicted in the following figure:

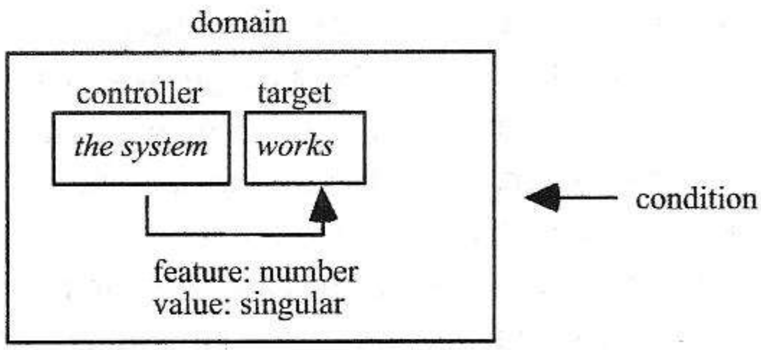


Figure 1: The framework of agreement terms (Corbett 2006:5)

From the generative point of view, agreement establishes a relation between two kinds of elements if they share certain grammatical features (Chomsky 2000, 2001; Kremers 2003). This operation consists of two elements: a probe and a goal[1]. The probe can enter into an agree-relation if it is active. This probe can be active when it has an unvalued feature like lacking φ-features (features of gender, person and number). Hence, it can have its features valued by probing for an active goal in its c-command domain that has the same matching features but valued (Chomsky 2000, 2001). The probe is the target that seeks for the φ-features, while the controller is the goal that bears

[1] A probe is a head that searches for a constituent (goal) within its c-command domain to agree with. C-command refers a structural relation between two constituents (X and Y) in which X c-commands its sister Y and any constituent Z that is contained within Y.

the φ-features. The φ-features are valued from the goal to the probe; they are morphologically realized on the probe. These features can be interpretable if they have a semantic interpretation; if they do not, they are uninterpretable. In English, they are interpretable on nouns, but they are uninterpretable on verbs. For example, the plural morpheme –s in *books* yields a different meaning in comparison with the singular *book*, so the φ-features on the noun are interpretable. However, the number feature on a verb does not have meaning (like *She plays* and *They play*.). This indicates that the φ-features of the verb *play* are uninterpretable.

In order to have a better understanding of the notion of agreement in generative grammar, in the next subsection, we review competing theories of word structure that discuss the relation between syntax and morphology; the extent to which syntax and morphology interact.

2.1.1 The relation between syntax and morphology

There is a great variety of theories on morphological inflection that result from a theoretical discussion of how morphology relates to the structures generated by the syntax. There have been two main views over the past few decades on how these two modules are related, with the key difference in whether morphology is pre-syntactic or post-syntactic. The first view is referred to as Lexicalism in which words are built in the lexicon by distinct mechanisms that are different from the mechanisms that create syntactic structure (Chomsky 1970, 1995; Lieber 1992; Lapointe 1980, 1981; Kiparsky 1982; Di Sciullo and Williams 1987). The second view is referred to as Distributed Morphology in which morphemes are not assigned in the lexicon. Instead morphemes are assigned to syntax and later spelled out by phonology (Halle and Marantz 1994, Harley and Noyer 1999, Embick and Halle 2005, Embick and Noyer 2007, Harley 2010).

Lexicalism is based on the assumption that word formation and phrase formation belong to two independent components of grammar and that there is a strict division of labor between them. Lexicalism comes in two varieties, strong and weak versions. Strong Lexicalism (Lapointe 1980, 1981; Kiparsky 1982; Di Sciullo and Williams 1987; Lieber 1992; Chomsky 1995) is the view that derivational and inflectional processes take place in the lexicon (see figure 2). A strongly lexicalist theory treats both inflectional and derivational forms as internally impenetrable to syntax. There is complete separation of morphology and syntax. Thus, word structure and syntactic structure have no direct access to each other. The only way they are related to each other is by the lexical insertion operation. This operation introduces the word forms with their associated feature structures into the syntactic structure.

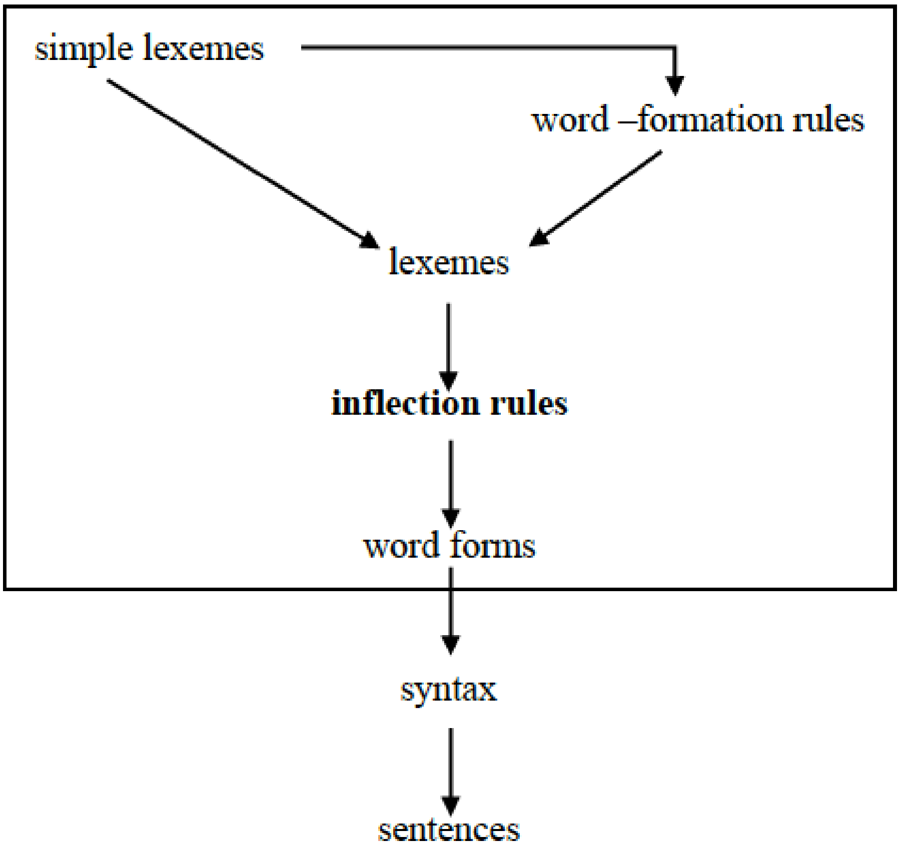


Figure 2: The strong lexicalist architecture

In contrast, weak Lexicalism (Chomsky 1970; Lapointe 1980, 1981) is the view that is based on a sharp distinction between word formation and inflection. It allows interaction between word structure and syntax only in the domain of inflectional morphology. Therefore, with respect to syntax, derivational morphology is treated as enclosed. However, inflectional morphology is allowed to be determined by the syntactic component. A word formation component produces complex words derivationally. It also produces stems (lexemes) that acquire the morpho-syntactic features relevant to their inflection by means of their place in the clause structure and their participation in syntactic relationships. The syntactic component can interact with the morphological component and applies inflectional rules to the lexeme. Accordingly, the appropriate inflected form of this lexeme is derived. A schema of this view is provided in figure 3 below.

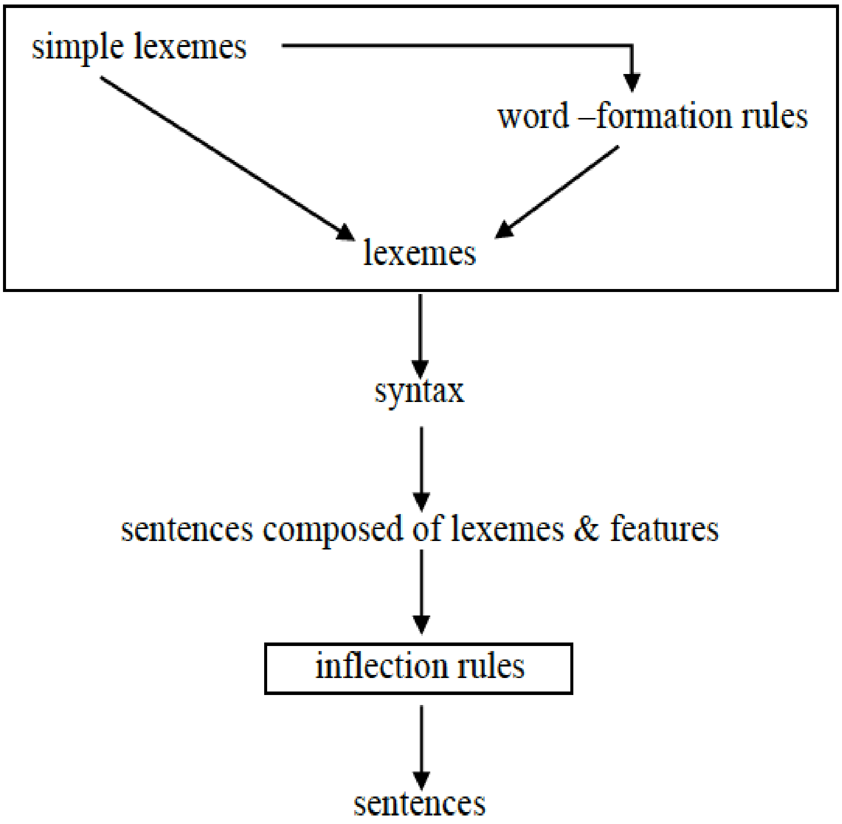


Figure 3: The weak lexicalist architecture

The second view that relates to the general standpoint in most current generative frameworks is that morphology (both derivational and inflectional) can interpret the output of syntactic structures. Thus morphology is entirely post-syntactic. This theory has been termed 'Distributed' Morphology' (DM) (Halle and Marantz 1993; Harley and Noyer 1999; Embick and Halle 2005; Embick and Noyer 2007). It adopts the view that the syntactic component constructs words and phrases alike. Since the mechanism that builds up the complex words are basically the same as the one that builds up syntactic structure, the interface between syntax and morphology is direct.

DM adopts the architecture of the grammar as sketched in figure 4, in which the syntax contains a set of rules that generate syntactic structures, and then these structures are subjected to further morphological operations that apply during mapping from the output of a syntactic derivation to the (input to) the phonology.

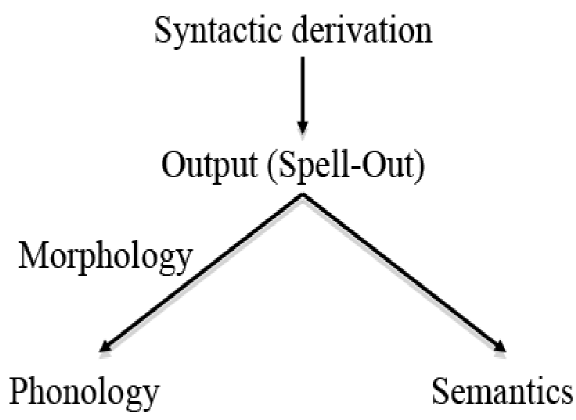


Figure 4: The architecture of grammar in DM (based on Embick and Noyer 2007)

To clarify the syntax/morphology interface, DM posits that in the grammar two types of morphemes can be found to serve as the terminals of both the syntactic derivation and word formation:

- Roots: These makeup members of the so-called 'lexical categories' like nouns, verb, etc. For example, the noun ox consists of the root *OX* and the verb hit consists of the root *HIT* that are sequences of complexes of phonetic features without any grammatical features.
- Abstract Morphemes: These are composed only of grammatical features (no phonetic features). They are functional categories of syntactic theory such as [PAST] or [PL]

(Embick and Noyer 2007).

For example, on the morpho-syntactic level, there are two kinds of elements: a root *WALK* and an abstract morpheme [PAST]. At this level, these two elements are combined into one abstract syntactic object without phonological content. After syntax, at the morpho-phonological level, phonological expressions (called Vocabulary Insertion) are added to the root and the abstract morpheme in a process called spell-out. In this process, Vocabulary Items (rules about where a phonological string or piece can be inserted) are added. If multiple morpho-syntactic features are realized in one Vocabulary Item, abstract morpho-syntactic morphemes are merged with the syntactic tree by a fusion rule, before the vocabulary item is applied. In cases when morpho-syntactic features are identified by vowel change instead of an additional morpheme, a zero suffix is inserted, before readjustment rules perform the necessary item-specific phonological operations (cf. Embick and Halle 2005). In the same vein, Halle and Marantz (1993) make a distinction between primary exponents (the addition of the affixes) and secondary exponents (the other changes to the stem). At analyzing IVs in English, Halle and Marantz state that the first step is to insert the primary exponents (the rules in (2)). That is, a morpho-syntactic node I, which results from the fusion of the syntactic nodes Tns (Tense) and Agr (Agreement), is spelled out by the following rules (Halle and Marantz 1993: 126):

2. I = the fusion of Tns and Agr

[+participle, +past] ↔ /-n/ / X + ______

where X = ^hew, go, beat,…

[+past] ↔ /-∅/ / Y + _____

where Y = beat, drive, bind, sing, …

[+past] ↔ /-t/ / Z + ______

where Z = dwell, buy, send, …

[+past] ↔ /-d/

[+participle] ↔ /-ing/

[3sg] ↔ /-z/

↔ /∅/

These rules show the competition between affixes for the spell-out of inflectional features, as they are disjunctively ordered. Therefore, this ordering will guarantee the blocking of the form *singed as a past tense form of the verb sing, as sing undergoes an earlier rule. This asserts the addition of a zero-affix as a spell-out of the past tense feature and blocks the insertion of the default rule *–ed*. The form sang only appears after the application of a second type of rule that is called readjustment rule. Halle and Marantz (1993: 128) give an example of such a readjustment rule that changes the vowel in the verb do in different morpho-syntactic environments:

3. Rime → /i/ / Y _____ [+past, -participle]

|

x

b. Rime → /Λ/ / Y _____ [+past, +participle]

| [-past, 3sg]

x

where Y- Rime = do

Again, it is important to notice that Rule (3-a) applies in the past tense and accounts for the form *did* only after the suffix –d has been added by one of the rules in (2). The same thing can be related to *done* and *does*.

Within the framework of DM, Yang (2002, 2005) presents the Rules and Competition theory. This theory describes a set of phonological rules to explain the English past-tense inflection and linguistic productivity in general. We will come back to Yang's work in the next section when we discuss morphological theories in language acquisition, as it is relevant for the current study.

To sum up, under the view of DM, syntax is the engine that combines abstract bundles of features of a word, while morphology is realizational. Morphology provides morphological content to syntactic structures already built. Consequently, morphology does not have an effect on how the syntactic structures are generated in the course of the derivation. In contrast to DM, Lexicalist theories regard the lexicon as a central component of language representation. Under the strong view, processes of both derivational and inflectional morphology occur in the lexicon, while the lexicalist weak view allows interaction between word structure and syntax in the domain of inflectional morphology. So, from the lexical perspective, morphology drives syntactic structures. Nevertheless, from DM perspective, syntax drives morphological structures.

After a long period of domination by generative grammar that constitutes prototypes of rules-only models for linguistic processing, connectionism succeeds to afford a different understanding of this processing. Connectionist approaches put forward specific assumptions all linguistic knowledge is learned and represented in an associative memory. Frequency is the key factor to establish associations among words in these approaches (Rumelhart and McClelland 1986; Bybee 1995). This different understanding has fueled the ongoing debates on the morphological acquisition in specific and the mental representation of language in general (Rumelhart and McClelland 1986; Pinker and Mehler 1988; Pinker and Prince 1988; Smolensky 1996; Bybee 1995; Seidenberg and Gonnerman 2000; Pinker and Ullman 2002, McClelland and Patterson 2002). The aim of the next section is to present the major tenets, strengths and shortcomings of models that play central roles in this debate.

2.2 The acquisition of inflection: theoretical approaches

A longstanding debate in linguistics and psycholinguistics relates to how linguistic information is processed by the human mind (Chomsky and Halle 1968; Rumelhart and McClelland 1986; Pinker 1999). The acquisition of English past tense morphology has become a battleground for this linguistic debate. In this debate, one question arises as to how linguistic knowledge, more specifically morphological knowledge, is mentally represented. In this respect, two different types of approaches can be distinguished: single and dual mechanism approaches of morphological processing (see figure 5). Single mechanism approaches posit no fundamental distinction between regular and irregular inflections, and contend that both are built via a single mechanism. Thus, these approaches hypothesize that all morphological processes are taken care of by one single mental mechanism – either a rule system or an associative system. Focusing on single mechanism approaches, followers of rule-based models assume that both RVs and IVs are generated by rules (Chomsky and Halle 1968; Halle and Mohanan 1985; Yang 2002). By contrast, supporters of storage-based models assert that all inflected words are stored within a single associative system (Rumelhart and McClelland 1986; Bybee 1995). Along with the dual mechanism approach, the core features of generative grammar and connectionism are combined: IVs are stored in the associative memory, while RVs are generated by rules. In the next section, I will provide more details about single and dual mechanism approaches.

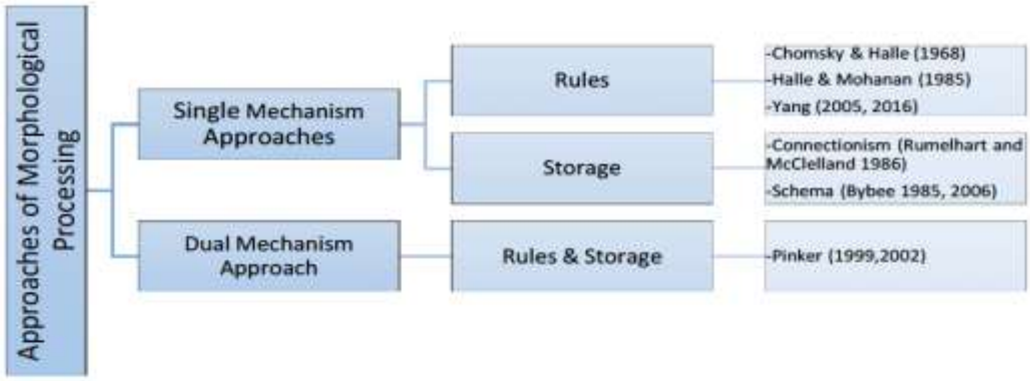


Figure 5: Theoretical approaches of morphological processing

2.2.1 Single mechanism approaches

2.2.1.1 Rules only

The traditional generative approaches (e.g. Chomsky and Halle 1968; Halle and Mohanan 1985) assume that the memory system only contains morphemes that are joined together by linguistic rules to form morphologically complex forms and phrases. Of specific relevance to the present study is the way Halle and Mohanan (1985) handle the formation of both English past tense verb RVs and IVs. RVs are generated by the default rule that adds a suffix *–ed* to the stem of the verb (e.g., *walk-walked*). For IVs, there is a number of morpho-phonological rules, which bonds verb stems and groups of verb stems to their related past and perfect forms. It is important to note that the morpho-phonological rules apply before the default rule, where the 'default condition' is analyzed as 'the elsewhere condition' (Kiparsky 1982). Accordingly, only forms that have not undergone the morpho-phonological rule processes are subject to the application of the default rule. This process prevents the production of over-generated forms like ranged and marks regular suffixation as the 'default' process.

Consider the following two examples that show how the proposed morpho-phonological rules in Halle and Mohanan (1985) generate IVs:

4. Lowering:

V → _[+low, -high]

Halle and Mohanan (1985:107) claim that the above-stated ablaut-lowering rule accounts for verbs such as *sit, begin, drink, sing, spit, ring, spring*, and a few more. These verbs have past tenses in which the stem-vowel is /æ/. Since the verbs all have an /I/ in the present tense, Halle and Mohanan propose the above rule that change the stem vowel /I/ to /æ/ in the past tense:

5. Backing Ablaut:

$$\begin{bmatrix} V \\ \langle -\text{high} \rangle_a \end{bmatrix} \longrightarrow \begin{bmatrix} +\text{back} \\ \langle +\text{round} \rangle_b \end{bmatrix}$$

If a, then b

Halle and Mohanan (1985)

The ablaut-backing rule presented above contains two parts. The first part of the rule accounts for verbs such as *cling, slink, spin*, etc. These verbs change their stem vowels from / I / to / Λ / as a result of backing. The second part of the rule accounts for verbs, such as *swear, wear, bear*, etc. The stems of these verbs contain the non-high vowels that are not only backed but also rounded to produce past tense forms such as *swore, wore, bore*, etc.

Hence, Halle and Mohanan attempt to account for the sub-regularities within the set of roughly 180 irregulars in the past forms in English. They suggest 10 morpho-phonological rules that are applied to stems of IVs that are stored in the mental lexicon.

Needless to say, these morpho-phonological rules are restricted only to specific lists of verb stems. But, it is common that certain irregular patterns are productive in the sense that they are extended to new stems on the basis of phonological similarity. These new stems, however, are not marked for a specific morpho-phonological rule such as the nonce-word *spling* that is created by researchers for linguistic experiments (Prasada and Pinker 1993; Xu and Pinker 1995; Albright and Hayes 2003). This nonce-word is likely to be inflected by adults and children as *splung* on the basis of its phonological similarity to verbs e.g., *cling, spring*, and *fling*. The Halle and Mohanan model does not predict such behavior, since the nonce-word *spling* is not stored in the lexicon and thus cannot be marked by the grammar to undergo a specific vowel alternation rule. Likewise, it is known that children and adults may produce certain irregularization instances by applying irregular patterns to already existing irregulars. For example, *bring-brang* may be produced in reference to *ring-rang* (Berko 1958; Marcus et al. 1992; Pinker 1999). Marcus et al. 1992 argue that such irregularization instances should not be produced if *bring* is marked in the grammar to undergo a specific rule that changes the stem *bring* into the past tense form *brought*. In this sense, thus, Halle and Mohanan (1985) disregard the productivity of such irregular patterns.

In the same respect, suggesting that only stems are stored in the mental lexicon and that all subsequent inflections are generated by rules runs counter to psycholinguistic findings that have been attested in many studies. For example, in one of the psycholinguistic experiments of speeded production tasks (Prasada, Pinker and Snyder 1990), it has been attested that irregular past tense forms display frequency effects. That is, stem frequency being equal, IVs with low frequency are produced (or comprehended) slower than IVs with high frequency. This speaks against the view that irregular past tense forms are produced by means of rules that are applied to stems. Instead, it may offer evidence for whole word storage (Say 2000). Yet, these frequency effects have not been attested for regular past tense forms. This may be indicative of inflection by the default rule that is assumed in the Halle and Mohanan model.

Inline with the rule-based models (Chomsky and Halle 1968; Halle and Mohanan 1985; Halle and Marantz 1994), Yang conjectures that the regulars are subject to the default rule. In contrast to the dual mechanism approach, Yang suggests that the irregulars are also subject to systematic rules, in lieu of retrieving them from memory. Hence, Yang (2002, 2005, 2015) establishes a fully rule-based account for the acquisition of inflected forms. In the tradition of DM, Yang suggests that the critical distinction is between being subject to a ‘more general’ rule or a ‘less general’ rule. The default rule is the most general rule. Therefore, the only distinction between default and non-default rules is that non-default rules are applied to specific contexts, whereas the default rule is not restricted to such contexts. So, the default rule is considered the most general rule. The task for a learner is to discover the default inflectional rules of the language and memorize that forms are subject to specific rules. Yang further argues that inflectional rules emerge in the learner as co-existing and competing hypotheses.

The learner has to decide for each irregular form whether the default rule or a specific rule applies (Yang 2002: 61).

For the formation of the English past tense, Yang proposes that there is the default rule that adds *-d* to the verbal stem of most verbs. In addition, he assumes that even IVs form their past tenses by using special rules that need to be memorized. If the learner knows only the forms *sing-sang* and *ring-rang*, it will try to build the following rule:

6. In case of /Xing/ change to /Xang/

But, when the learner 's vocabulary grows, he/she will face more exceptions such as *bring- brought, swing- swung* and *wing-winged*. At this point, the learner determines that the rule is limited in scope, as it just applies in a particular set of verbs. Thus, the rule will be stored with the additional information. Following an original proposition introduced by Anderson (1974), Yang calls this type of rule a morpho-lexical rule (a rule with limited productivity that merely applies in a limited set of the forms that one would presume given its structural description). These morpho-lexical rules are arranged according to 'The Elsewhere Condition', in that a more specific insertion context will take priority over a less specific one. Thus, the morpho-lexical rules are ordered before the default (productive) rule that has to come last. In (7), two rules that are associated with a set of verbs are pictured:

7. {feed, shoot,….} → R Vowel Shortening

   {bring, think,…} → R-t suffixation & Rime→ a

(Yang 2002: 64)

Verbs belonging to the class of *feed, shoot*, etc., can form their past tenses through vowel shortening. Verbs belonging to the class of bring form their past tenses by means of *–t* suffixation and a change in the stem-vowel. If a verb does not fit to either of these classes, then it will build its past tense by using the default rule. It is assumed that the child can easily take up the particular phonological changes that compose the possible markings of the past tense in IVs (Yang 2002). What supports this claim is that young children barely make any mistakes in the formation of the past tense (about 90%

correct, according to an empirical study in Marcus et al. 1992). Even if they make mistakes, they are only as a result of overgeneralization of the default rule, not in the phonological changes.

Yang (2002:71) presents the schema in figure 6 to show how his model works. At the first stage, when presented with a past tense form, the child will compose the root (X). Then, the child should decide to which class of verbs this root (X) belongs in order to choose the appropriate rule for its past form. The value P (X in S) refers to the chance that the child assigns the verb X to class S. This value determines the choice of the appropriate rule. Now, if the child determines that X does not belong to any irregular class S, the default rule will step in to form its past tense. Then, this default form will be checked to see whether there is a match with the input X past. If there is a match, the value P (X in S) will be decreased. But, in case the child determines that X does belong to S, then it has to make the second choice: whether to apply rule R or not. The probability value P(R) (the chance that the child applies R to X) determines this choice. Once again, if the child prefers not to apply R, then it will choose the default rule. If this form will match with the input, the value P(R) will be lowered. But, in case that the child decides to apply rule R, here there will be two options: either this form will match the input and this increases the values P (X in S) and P(R) or it does not match and again this results in a decrease of P(R).

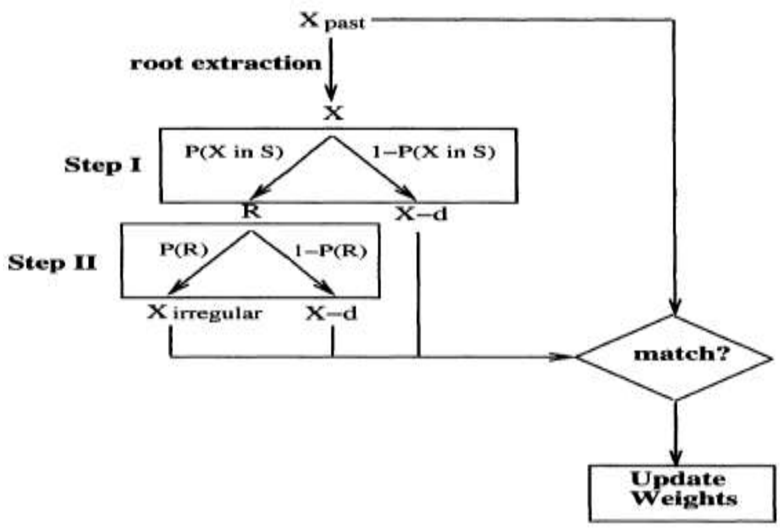


Figure 6: The schema of Yang (2002: 71) for learning IVs by rule competition

As P(R) is increased with every occurrence in the input of a verb from class S, this encourages Yang to make two precise predictions on the acquisition of IVs and the role of frequency in the input. First, verbs (within the same class) with greater frequency will be learnt faster than verbs with lesser frequency. Second, in case that there are two verbs with the same frequency but from different classes, the verb that derives from a class with greater frequency will be learnt quicker than the verb that derives from a class with lesser frequency.

To prove that these predictions are correct, Yang (using the corpus from Marcus et al. (1992)) calculates the value of the correct usage of a particular verb X by dividing the total number of correct past tenses of X by the total number of past tense of X in for each irregular verb. He concludes that children are enormously good at acquiring the past tenses, as he finds out that the average correct use over the four children in the corpus is 89,9%. In addition, within the same class, it is proven that verbs with greater frequency are less prone to overgeneralization than the lesser frequency verbs. For example, in table 1, Yang (2002: 82) gives the following results the verb class characterized by no suffix and no change [Ø + no change]:

Table 1: The frequency of the verb class characterized by no suffix and no change (after Yang 2002: 79)

Verb	**Correct use in corpus**	**Verb frequency in corpus**
put	239/251 = 95.2%	2248
hit	79/87 = 90.8%	66
hurt	58/67= 86.6%	25
cut	32/45= 71.1%	21
The average correct use	89,9%	

It is clear that the verb with higher frequency (e.g., *put*) carry the higher percentage of correctness (95.2%). Yang displays that the same prediction holds for the other verb classes.

In the second prediction, in case we have verbs with the same frequency but from classes that differ in frequency, the verbs that belong to the more frequent class will be learnt quicker than the verbs that belong to the less frequent class. To prove this, Yang compares the verbs *hurt* and *cut* from the verb class [Ø + no change] with the verbs *know* and *throw* from the class that forms the past tense only with a change of vowel of the stem and he gets the following results in table 2 below:

Table 2: The frequency of the two verb classes: with no suffix and no change and with a change of vowel of the stem only (after Yang 2002: 81)

Verb	**Verb class**	**Verb frequency in corpus**	**Correct use in corpus**
hurt, cut	[-Ø & No Change]	hurt (25), cut (21)	80.4%
know, throw	[-Ø & Rime → u]	know (58), throw (31)	49.1%

We see that in spite of the higher frequencies of the verbs *know* and *throw*, the verbs *hurt* and *cut* are less prone to overgeneralization. This is because the verb class to which *hurt* and *cut* belong also contains very high frequency verbs such as *hit, let, set,* and *put*. This high frequency class leads us to imply that the value of P(R) for this class will be extremely high and accordingly P (X in S) for verbs belonging to this class is also high although the frequencies of these verbs are relatively low.

From the discussion above, one may conclude that the first empirical result is not surprising at all, as it may be intuitively detectable that the frequency in the input of a particular irregular past tense correlates with a number of errors that is made in these verbs. Nevertheless, the second result cannot be easily accounted for, as the verbs that belong to the more frequent class will be learnt quicker than the verbs that belong to the less frequent class.

The way that Yang (2002) explains regularization processes is of particular relevance to the present study. He assumes that these processes are described through probabilistic strategies. During the process of language acquisition, the learner uses the probabilistically most advantageous rule, leading to overapplication of the regular default rule to forms that require specific (non-default) rules. Yang argues that regularization instances are thus misapplied phonological rules. Yet, Embick and

Marantz (2005: 245) propose that consulting a rule requires language learners to depend on their memory. The language learner must remember which stem form is located on which list. For example, for *ring*, a learner needs to know that there is a -Ø realization of the past tense, and that *ring* is on the specific list of verbs that appears with -Ø. One may argue that the failure to apply the appropriate phonological rules involves a memory lapse. This means that either the verb has not been stored on the suitable list or the stored verb has not been correctly retrieved from its list.

2.2.1.2 Storage only

A different approach to morphology – and to language as a whole – comes from the connectionist paradigm (Rumelhart and McClelland 1986; MacWhinney and Leinbach 1991; Bybee 1995, 2001; Joanisse and Seidenberg 1999; Plunkett and Juola 1999; Moscoso del Prado Martín et al. 2004). The central connectionist principle is that interconnected networks of uniform units can explain human language production. These units and connections mimic the functions of neurons and synapses respectively in the brain of a human being. The connectionist model is a very different approach from that of the rule-based models (Chomsky and Halle 1968; Halle and Mohanan 1985; Halle and Marantz 1994), as it does not use explicit rules at all. Rumelhart and McClelland (1986: 217) claim:

Instead, we suggest that the mechanisms that process language and make judgments of grammaticality are constructed in such a way that their performance is characterizable by rules, but that the rules themselves are not written in explicit form anywhere in the mechanism (Rumelhart and McClelland 1986: 217).

Rumelhart and McClelland (1986) are one of pioneers in this field of work. In their model, the pattern associator is used to model English past tense formation. This pattern associator mainly consists of two elements: an input layer (a pool of input units representing the verb base form) and an output layer (a pool of output units representing the past-tense forms) (see figure 7). Each input unit is connected to each output unit. The more often certain input (such as features of the sounds in a verb stem) occurs, the stronger the connections between it and its output will be. General cognitive mechanisms work on the stored mappings of forms and meanings, recognize common patterns and form analogies to similar cases (Rumelhart and McClelland 1986; Plunkett and Marchman 1991, 1993; Bybee 1995; McClelland and Patterson 2002; Cameron-Faulkner et al. 2003; Tomasello 2003, 2009). Using this learning mechanism, learners form analogies about almost any part of their world.

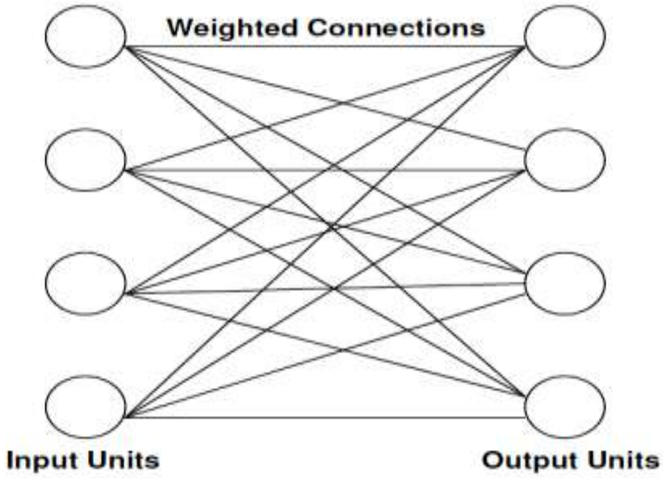


Figure 7: A simplified representation of the Rumelhart-McClelland model of past-tense inflection (Rumelhart and McClelland 1986)

From the connectionist perspectives, RVs and IVs are equally extracted from the input and stored in a single mechanism. For regular English past tense form, the network duplicates the features of the stem to the past-tense form and adds the suffix –*ed* e.g., *play-played*. For irregular English past tense form, 'the network uses the same connection-based knowledge that produces default forms and additionally taps into specific connections activated by the particular properties of keep to produce the vowel adjustment' (McClelland and Patterson 2002, cited in Fleischhauer 2013), such as *ring-rang*. With an attempt to maximize commonality between stored forms and a new form in the network, connectionists aim to include RVs and IVs, as well as instances of (ir)regularization, into one explanatory model in terms of analogy.

Nevertheless, the Rumelhart and McClelland model (1986) has serious generalization problems with regulars (Pinker and Prince 1988). The model provides incorrect responses including strangely inflected forms (*squat-squakt, mail-membled, tour-toureder, mate-maded)*, no change at all (*hug-hug, smoke-smoke*), double markings (*type-typeded, step-steppeded, snap-snappeded, map-mappeded, drip-drippeded*), and incorrect vowel/consonant-changes (*shape-shipt, sip-sept, slip-slept, brown-brawned*). As Pinker and Prince note, these production patterns definitely do not reflect the production patterns of human beings. Yet, the connectionist model succeeds to generalize most of the irregular patterns properly (Prasada and Pinker 1993). The model is able to produce the past tense forms of phonologically diverse verbs e.g., *slept*, *cut* and *flung* as past forms of *sleep*, *cut*, and *fling* respectively. This refers to its possibility to make complicated generalizations on the basis of phonological similarity to forms previously stored in the associative memory.

To sum up, single mechanism approaches posit no essential distinction between regular and irregular inflections. These approaches hypothesize that all morphological processes are built via one single mental mechanism (rules or storage). Followers of rule-based models adopt the view that both RVs and IVs are generated by rules (Chomsky and Halle 1968; Halle and Mohanan 1985). By contrast, proponents of associative models claim that all inflected words are stored within a single associative system (Rumelhart and McClelland 1986; Smolensky 1996; Bybee 1995; Seidenberg and Gonnerman 2000; McClelland and Patterson 2002). In an attempt to form a midway between entirely rule dependent models and entirely associative models, the dual mechanism approach defends the necessity of two separate mechanisms for language processing: rules for RVs and storage for IVs (Pinker, 1991, 1999; Marcus et al. 1992; Pinker and Ullman, 2002 among others). More details for this approach in the next section.

2.2.2 Dual mechanism approaches

Pinker (1999) characterizes English past tense morphology as a prime example to understand two distinct mechanisms that lie at the basis of the human language faculty:

The premise of this book is that there are two tricks, words and rules. They work by different principles, are learned and used in different ways, and may even reside in different parts of the brain (Pinker 1999: 2).

Therefore, every human being is endowed with a set of linguistic rules and a lexicon (Pinker and Prince 1988; Prasada and Pinker 1993; Pinker 1998; Pinker and Ullman

2002). Pinker's ideas on language and about the actual origin of those rules can be defined as innatist, as he claims that:

Inside everyone's head there is a finite algorithm with the ability to generate an infinite number of potential sentences, each corresponding to a distinct thought. The meaning of a sentence is computed from the meanings of the individual words and the way they are arranged (Pinker 1998: 3).

Pinker (1998: 223) focuses on RVs and IVs as a means to show that words and rules are the "ingredients" of language. These forms are considered to be the ideal testing ground for a number of reasons. Both types are equated for length and complexity (being single words), for grammatical properties (being nonfinite forms, with identical syntactic privileges), and meaning (both expressing the pastness of an event or state)". He affirms that regulars and irregulars are processed by different mechanisms: regulars are formed by rules, while irregulars are stored in the mental lexicon. So, a combination of the main features of the previous two theories are proposed:

- Generative grammar - productive rules
- Connectionist - an associative storage facility

In the same vein, Pinker and Ullman (2002) claim:

The regular–irregular distinction is an epiphenomenon of the design of the human language faculty, in particular, the distinction between lexicon and grammar made in most traditional theories of language (Pinker and Ullman 2002: 456).

They define the lexicon as a subdivision of memory that includes all arbitrary pairs of sound and meaning that refer to the morphemes and simple words of a language, while the grammar is a productive system that assembles morphemes and simple words into complex words, phrases and sentences. They also claim IVs, like other words, are acquired and stored, but with a grammatical feature like [PAST] merged into their lexical entries. RVs, by contrast, can be productive, so they can be generated by a rule, just like phrases and sentences. A stored inflected form of an irregular verb may block the application of the rule to that verb. When there is no stored irregular inflected form for a particular stem, the default affix of a regular form is attached to that stem; so this default affix applies for any unsorted item as in the following figure:

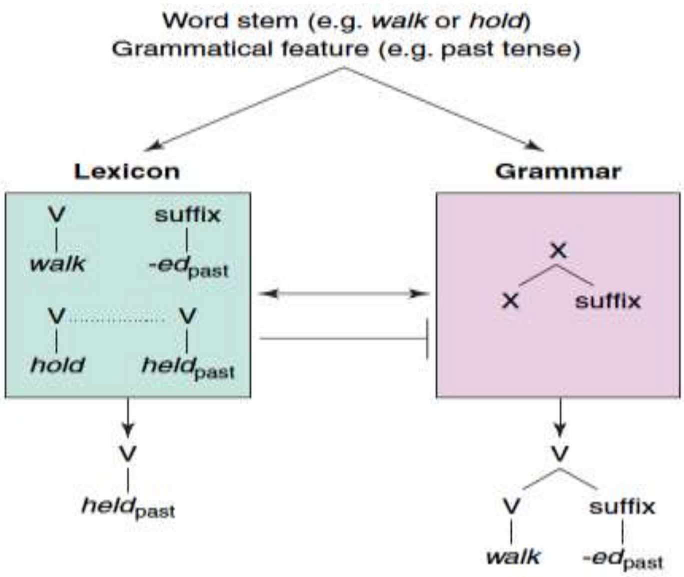


Figure 8: Simplified illustration of the Words-and-Rules theory (based on Pinker and Ullman 2002: 457)

Hence, when a word is inflected, it (in parallel) accesses the lexicon and the grammar. If an inflected form of a verb is found in the mental lexicon (as with irregulars e.g., *held*), it will be successfully retrieved. Due to this match, the operation of the grammatical suffixation process is blocked preventing the generation of a form like *holded*. If no inflected form is matched, then the grammatical processor searches for the suitable affix to be attached to the stem and generates a regular form. In this case, thus, the default rule is applied e.g., *walk*-*walked*.

For a better understanding of the implications of this theory, we need to know what is meant by the process of morphological blocking. This process is based on the generative dogma that it is impossible to have two forms expressing the exact same idea within an individual's linguistic repertoire at a given point in time. Don et al. (1994) claim that this view refers to the phenomenon where the occurrence of one word blocks the possibility of generating another word with the identical meaning e.g., *ran*-*runned*.

Pinker (1984) outlines this concept as the ‘unique entry principle’. This principle refers to “the constraint that no cell in a paradigm may be filled with more than one affix” (Pinker 1984: 177). Hence, when an irregular past tense is stored in the lexicon, it prevents the rule from applying:

If a word can provide its own past tense from memory, the regular rule is blocked: that is why adults, who know broke, do not say breaked. Elsewhere (by default), the rule applies: that is why children can generate ricked and adults can generate moshed, even if they have never had a prior opportunity to memorize either one (Pinker 1998: 223).

Accordingly, whenever there is the possibility of a successful retrieval of an irregular form e.g., *slept*, this is said to result in the sending of a constant signal to the rule-system, which in turn blocks the application of the default rule and therefore prevents the occurrence of RFs e.g., *sleeped*. Nevertheless, if a memorized form cannot be retrieved because of various factors (e.g., no memory entry of irregulars as in the case of nonce-forms (e.g., *ploamph*) or the weakness of the memory traces of irregulars as in the case of newly acquired or very low frequency forms), the regular rule applies. The regularization form e.g., *sleeped* instead of *slept* is called a “blocking-and-retrieval-failure” (Marcus et al. 1992; Marcus et al. 1995).

Focusing on verbs’ sensitivity to frequency, let’s compare between RVs and IVs. As RVs are produced by means of a rule, regular inflection is supposedly frequency insensitive. Pinker (1999) emphasizes this hypothesis stating that a regular form turns to have a high type frequency. The regular past tense inflection of -*ed* applies to 86% of the 1000 most common verbs (Pinker 1999). Another evidence for the regulars’ insensitivity to frequency comes from plurality in German. The plural marker -*s* (as in *Kinos* (cinema.PL)) only applies to a small number of words. There are other more common plural markers e.g., -*e* (as in *Stifte* (pen.PL)), -er (as in *Bilder* (picture.PL)), -*n* (as in *Vasen* (vase.PL)). Nonetheless, Clahsen (1999) and Marcus et al. 1995 present experimental evidence that even in this case of ‘minority inflection’, the plural marker -*s* is still considered as the default marking. It is generalized to novel and derived words where no stored irregular form is available in the mental lexicon (e.g. plural *Kachs* rather than *Käche* or *Kacher*).

However, the situation is different with IVs that are memorized as lexical items in the mental lexicon. Thus, irregular inflection should be sensitive of frequency. In order to store IVs as wholes in memory, these forms should be frequent enough in language use to warrant such storage. Hence, in the process of learning, there is a certain threshold for complex forms to be stored in memory. If the frequency of the certain irregular form is above this threshold, then this form will be stored in the lexicon. If not, for whatever reason, the irregular form will regularize, as the default rule will be used. This is because the use of the default rule takes some effort every time it supposes to be produced. This effort is lesser in case the form can be retrieved from memory on the whole. In this regard, Pinker (1999) exposes two predictions: first, high frequent irregulars are retrieved faster than low frequency irregulars. Hence, low frequency irregulars are regularized more often than high frequency irregulars. Only words with high frequency can remain irregular without being subsumed by the regular rule. Thus, irregulars tend to have high frequency, and high frequency verbs tend to be irregular. Second, no significant difference in retrieval rates between high and low frequency RVs. There are some empirical findings that support these predictions. The top ten frequent verbs in English, *be, have, do, go, say, can, will, see, take* and *get*, are all irregular (Pinker 1999; Lieberman et al. 2007). Similarly, KuCera and Francis (1967) exhibit that the most frequent verbs of English are all irregular, while verbs with lower frequencies are all regular. They claim that of the top 30 most frequent English verbs in the past form 22 are irregular. Additional empirical evidence originates in diachronic change: IVs that in the course of history for whatever reason become less frequent get regular inflection, several examples such as *to chide*, *to glide* and to gripe have become regular during the course of history (Pinker 1999: 69).

More findings from various empirical studies of the morphological acquisition with/against single-dual mechanism approaches will be produced in the next subsection. The main focus will be on studies of (ir)regularization processes that is particularly relevant to work in this thesis.

2.2.3 Findings from empirical studies of (ir)regularization

From the dual mechanism perspective, IVs should be frequency-sensitive because they are stored in the associative memory. So, IVs with high frequency are more prone to remain irregular over time due to their strong representations in the associative memory that make them easier to be accessed. IVs with low frequency, however, are more prone to regularization processes due to verbs' weak representations in the associative memory that make them harder to be retrieved. Yet, RVs are rule-produced and hence are frequency-insensitive. Consequently, this approach predicts that language changes are in principle unidirectional, towards regularization only. Some empirical studies support this hypothesis (Marcus et al.1992; Lieberman et al. 2007; Michel et al. 2011 among other). For example, Lieberman et al. (2007) assert that frequency plays a central role in regularization processes revealing that the number of IVs has gradually declined over the past centuries. But, verbal changes in the other direction have been observed as well (Peters 2004; Nübling 2000 and Fertig 2013). According to single mechanism views, all morphological processes are taken care of by one single mental mechanism; rules or storage (Chomsky and Halle 1968; Halle and Mohanan 1985; Rumelhart and McClelland 1986; Bybee 1995, yang 2002; Peters 2004; and Fertig 2013). Consequently, followers of these approaches predict that language changes are bidirectional. They suggest that similar mechanisms underlie the production of both regularization and irregularization. Fertig (2013) claims that:

Regularization may be more common historically than irregularization, but irregularizations occur much more often than many linguists seem to realize, and it may be that – at least at some points in a language's history – attested variation is just as likely to reflect irregularization as regularization (Fertig 2013*: 92*).

Lieberman et al. (2007), have recently examined the historical trend towards regularization of English IVs. They calculate the regularization rate of 177 verbs from Old English to the present day, relative to their frequency of occurrence. They aim at relating the evolution of language to evolution in the hard sciences, such as genetics and nuclear physics. Table 3 below displays the 177 Old English IVs that are compiled for Lieberman et al.'s study. These verbs are divided into six bins according to frequency of occurrence and arranged in alphabetical order within each bin. RFs are identified in red. Over the last 1200 years, they find that of these 177 irregulars, 145 remained irregular in Middle English, and 98 are still irregular in Modern English. They draw a conclusion that IVs with low frequency are more quickly regularized, while IVs with high frequency remained that way much longer.

Table 3: 177 Old English IVs in the study of Lieberman et al. (2007: 17). Verbs that have regularized are indicated in red

Frequency	Verbs	% Reg	Half Life
$10^{-1} - 1$	be, have	0	38,800
$10^{-2} - 10^{-1}$	come, do, find, get, give, go, know, say, see, take, think	0	14,400
$10^{-3} - 10^{-2}$	begin, break, bring, buy, choose, draw, drink, drive eat, fall, fight, forget, grow, hang, help, hold, leave, let, lie lose, reach, rise, run, seek, set, shake, sit, sleep, speak, stand, teach, throw, understand, walk, win, work, write	10	5400
$10^{-4} - 10^{-3}$	arise, bake, bear, beat, bind, bite, blow, bow, burn, burst, carve, chew, climb, cling, creep, dare, dig, drag, flee float, flow, fly, fold, freeze, grind, leap, lend, lock, melt, reckon ride, rush, shape, shine, shoot, shrink, sigh, sing, sink slide, slip, smoke, spin, spring, starve, steal, step, stretch, strike, stroke, suck, swallow, swear, sweep, swim, swing tear, wake, wash, weave, weep, weigh, wind, yell, yield	43	2000
$10^{-5} - 10^{-4}$	bark, bellow, bid, blend, braid, brew, cleave, cringe crow, dive, drip, fare, fret, glide, gnaw, grip, heave knead, low, milk, mourn, mow, prescribe, redden, reek, row, scrape, seethe, shear, shed, shove, slay, slit, smite sow, span, spurn, sting, stink, strew, stride, swell, tread, uproot, wade, warp, wax, wield, wring, writhe	72	700
$10^{-6} - 10^{-5}$	bide, chide, delve, flay, hew, rue, shrive, slink, snip, spew, sup, wreak	91	300

In the highest frequency bin (10^{-1}–1), there are 2 IVs, be and have in the present day. In the second frequency bin (10^{-2}–10^{-1}), 11 IVs are found. All are still irregular in the present day. Thus, none of IVs in the two highest frequency bins have been regularized over the passing of time. In the third frequency bin (10^{-3}–10^{-2}), 37 irregulars are attested. 4 of them, help, reach, walk, and work, are regular in the present day. In the fourth frequency bin (10^{-4}–10^{-3}), there are 65 irregulars. 28 of them undergo regularization process. In the fifth frequency range (10^{-5}–10^{-4}), 50 irregulars are attested; 38 of them have been regularized over the passing of time. Finally, in the lowest frequency bin (10^{-6}–10^{-5}), 12 irregulars decline to only 1 (slink) in the present day; which means that 90% of these verbs have been regularized over the past 1,200 years. The regularization process in this study is represented as being not constant in this study. Lieberman et al. (2007:1) state that 'a verb that is 100 times less frequent regularizes 10 times as fast'. Hence, less frequent irregulars are regularized more quickly than more frequent irregulars. A support for this claim comes from the fact that the ten most common verbs are all irregular (*be, have, do, go, say, can, will, see, take, get*), though less than 3% of modern verbs are irregular (Lieberman et al. 2007: 2). Based on their findings of the verbs, Lieberman et al. make a prediction about the future of English irregulars claiming that in the year 2500 only 83 of the 177 verbs will remain irregular.

The study by Lieberman et al. (2007) receives a wide discussion in both scholarly and popular spots. Three scholarly studies are mentioned here. We start with the study of Carroll et al. (2012) who examine verb regularization over time with parallel data from a closely related language (German) using the same procedures in the study of Lieberman et al. The findings of the German data exhibit that the rate of verb regularization in the history of German is obviously different from the one in the history of English. To draw a comparison, table 4 below illustrates the regularization rate for English in Lieberman et al.'s study and the regularization rate for German in Carroll et al.'s study in the six bins. The regularization rate in the English data is the highest (91%) in the lowest frequency verbs (bin 6) that is very close to the regularization rate for German in bin 6 (83%). However, large discrepancies appear between the English bins 4 and 5 (43% and 72% respectively) and the German bins 4 and 5 (8.2% and 37.5% respectively). Hence, Carroll et al. 's findings from the German dataset clearly underscore that there is no universal rate of change, even for verbs within the close West Germanic family.

Table 4: Comparison of the regularization rates for English in Lieberman et al.'s study and for German in Carroll et al.'s study in the six bins (after Carroll et al. 2012: 162)

Bin number	**English**	**German**
1	0	0
2	0	0
3	20	2.1
4	43	8.2
5	72	37.5
6	91	83.3

De Clerck and Vanopstal (2015), in another study, argue that Lieberman et al. do not take into account that their list of IVs is incomplete. For example, this list does not contain IVs that were added in Middle English e.g., *dream* and *spill*. In addition, it does not include doublet verbs whose past forms allow both the regular *–ed* and the irregular *–t* in present day English e.g., *burnt/burned, leapt/leaped, dove/dived*. In line with previous research (Lieberman et al. 2007), De Clerck and Vanopstal explore the relationship between frequency of occurrence and regularity in the different varieties of English in a synchronic snapshot. Their data is drawn from the Corpus of Global Web-Based English (GloWbE). They examine -*ed*/-*t* preferences in 11 doublet verbs namely *burn, dream, dwell, kneel, lean, leap, learn, smell, spell, spill* and *spoil.* The results of their corpus study, unlike Lieberman et al.'s study, provide no support for the influence of frequency on regularity: high-frequency verbs do not show higher -*t* percentages and vice versa. In addition, De Clerck and Vanopstal examine the impact of vowel change on the retention of regularization of these verbs. Some of these verbs have no vowel change between IVs and their RFs e.g., *learned/learnt* and *spoiled/spoilt*, others undergo vowel change as in *leaped*/*leapt* and *lighted*/*lit*. Table 5 provides an overview of the selected verbs that are divided into two groups: *no vowel change group* and *vowel change group*. These verbs are arranged from high to low according to their preferences to regularization; the verbs in red show a preference of more than 50 %.

Table 5: Preferences of 11 IVs for RFs in no vowel change and vowel change groups (after De Clerck and Vanopstal 2015). The verbs that show a preference of more than 50 % are in specified red.

No Vowel Change	Vowel Change
learn	**kneel**
spill	**leap**
spell	dream
smell	lean
spoil	
burn	
dwell	

Table 5 displays that most of the verbs with no vowel change have preferences for RFs (i.e. *learn, spill, spell, smell, spoil* and *burn*). However, only half of the verbs with vowel change display preferences for RFs (i.e. *kneel* and *leap*). Hence, their study reveals that there is a link between salience of vowel change and regularization process: verbs with no vowel change are regularized more often than verbs with vowel change.

Finally, as mentioned before, Fertig (2013) rejects to the one-sided of Lieberman et al.'s view that looks only at morphological changes in the direction of regularization. Instead, he argues that verbal changes are bidirectional, both towards regularization and irregularization. Several of originally RVs have become irregular in English (e.g. *sneak-snuck, drag-drug, hang-hung, string-strung, stick-stuck, strike-struck, dig-dug, wear-wore, wake-woke, light-lit, ring-rang, catch-caught, kneel-knelt, make-made, cost-cost*). Fertig (2013) claims that irregular forms (IFs) are necessary to a balanced understanding of the history of English irregulars. He also claims that the influence of word frequency is different for regularization than for irregularization. He asserts that there is a correlation between low frequency and regularization. Also, he uncovers a correlation of high frequency with irregularization claiming that IVs with high frequency are irregularized more often than IVs with low frequency.

Now, let's focus on some empirical studies of irregularization processes. Xu and Pinker (1995) analyze instances of irregularization in 20,000 regular past and perfect uses from 9 children in the CHILDES database (MacWhinney 2000). They find that the irregularization rate is only 0.2% of the opportunities. From the dual mechanism perspective, the irregularization rate should be rare compared to the regularization rate; 4.2% (Marcus et al. 1992). A lower rate of irregularization compared to regularization in the dual mechanism view is in conflict with predictions by single-system approaches. In the dual mechanism approach, verbal changes occur mostly unidirectionally, in that regular inflection is applied to IVs but not the other way round. This is due to the hypothesis that RVs and IVs are processed via distinct mechanisms – a rule-based system for RVs and an associative system for IVs. However, single-system approaches predict that verbal changes are bidirectional, towards both regularization and irregularization. These approaches hypothesize that RVs and IVs are processed via one single mental mechanism – either a rule-based system or an associative system. Accordingly, they propose that similar mechanisms underlie the production of both regularization and irregularization, hence regularization and irregularization are expected be at similar rates. In addition, both dual and connectionist models predict that instances of irregularization display phonological neighbourhood effects due to their hypothesized full-form storage of irregulars in the associative memory. Nevertheless, in connectionist models, irregularization rates are between 3.2% and 23.5% (Rumelhart and McClelland 1986; Plunkett and Marchman 1991; Sproat 1992; Plunkett and Marchman 1993) that are much higher than the irregularization rate manifested in the study of Xu and Pinker (1995); only 0.2%.

In the same vein, Peters (2004, 2009) suggests bidirectional modeling of the evolution of English verb morphology arguing that evolutionary expectations of unidirectional modeling are generally overstated. Peters (2004: 540) states that historically there are instances of English verbs that add irregular parts to what have been previously RVs e.g., *thrive-throve, light-lit, ring-rang, catch-caught, kneel-knelt* and *make-made*. Some instances of irregularization are regionally conditioned. For instance, in British English, RVs *saw* and *sew* acquire irregular perfect forms *sawn* and *sewn*, although the -*ed* forms are stronger in American English (Peters 2004: 487). Using data from the ICE-corpora for Australian, New Zealand and British English, Peters (2009) tries to examine irregularization processes. His analysis of the selected data shows that people in Australia and New Zealand are more likely to reduce 3-part verbs (as with *sing-sang-sung, ring-rang-rung, drink-drank-drunk* etc.) into 2-part verbs (as with *cling-clung-clung, fling-flung-flung, slink-slunk-slunk* etc.). In addition, there are preferences for the use of -*t* suffix for past and perfect forms of verbs that allow for both RVs and IVs e.g., *burnt*/*burned*. Thus, he suggests that (ir)regularization of English verb morphology can be conditioned by time and place. Accordingly, Peters

(2009: 27) tries to model and analyze the directions of change in English verbs. He displays a complex model of verbal movements that allows for both regularization and irregularization for verbs with either two or three contrasting parts, as shown in the table below

Table 6: Modeling the directions of change in English verbs (Peters 2009: 27)

⟨⟨⟨ Regularization 2-part verbs		Irregularization ⟩⟩⟩ 3-part verbs	
(A) -*ed* (pt and pp)	(B) other past forms for pt and pp	(C) pt as -*ed* plus irreg. pp	(D) irreg. pt and irreg. pp
earn/earned	*build/built*	*prove/proved/proven*	*begin/began/begun*
help/helped	*bring/brought*	*sew/sewed/sewn*	*drive/drove/driven*
like/liked	*fling/flung*	*show/showed/shown*	*bite/bit/bitten*
melt/melted	*spin/spun*	*sow/sowed/sown*	*take/took/taken*

In this model, Peters sets verbs with two or three contrasting parts into four groups:

- Group A: on the extreme left hand side, this group includes RVs or verbs that have been reached the evolutionary target of regularization in Lieberman et al.'s (2007) terms. It covers 2-part verbs of -*ed* paradigm in which both past tense and past participle have the same form.
- Group B: this group involves 2-part verbs such as *cling-clung-clung*. Historically and in present day English, there is a tendency for reducing contrasting verb parts from 3 (i/a/u paradigm such as *sing/sang/sung*) to 2 (i/u/u paradigm such as *cling-clung-clung*) (Jespersen1965; Bybee 1982, Peters 2009). Peters (2009: 26) states that '[f]ounding members of the i/u paradigm are earlier refugees from the *sing/sang/sung* group, such as *cling, sling, slink, sting, stink, swing, wring'*. The i/u paradigm is also gaining in strength through adding new members such as *sneak-snuck* and *drag-drug*. These new members end in velar stops rather than velar nasals and have different vowel stems. Peters claims that the reduction of a verb's parts to 2 in this group puts it on a par with the -*ed* paradigm that has the same form for both past tense and past participle.
- Group C: here, verbs are a mix: they are strictly 3 part-verbs, but they are regular by their dental past tense and irregular in their past participle form.

Peters (2009: 27) states that '*sow* is a case of incomplete regularization, while *sew* (like *saw*) is a case of incomplete irregularization'.

- Group D: this group contains 3-part verbs: present ≠ past ≠ past participle such as *ring/rang/rung, drive/drove/driven, take/took/taken.* Together they generally display the attraction of the 2-part verb paradigm rather than regularization to the *-ed* paradigm.

The model of the bidirectional changes in English verbs, as mentioned before, comprise only verbs with two or three contrasting parts. It does not provide for one part verbs e.g., *put-put-put, wed-wed-wed*. This group generally displays no sign of regularization. Peters (2004: 574–5) claims that the general resistance of this group to regularization may be due to their high frequency and their phonology (short vowels before a dental consonant).

It is essential to take into consideration that much of the single-dual mechanism debate of language processing revolves around the mental representation of L1 knowledge. Hence, the principles of single and dual mechanism approaches are developed solely on the basis of findings based on monolingual individuals. Correspondingly, the main focus of many scientific studies of this processing, the ones mentioned above included, is on L1 acquisition and processing. Some studies also focus on L2 acquisition (e.g., Clahsen 1995; Beck 1997; Zobl 1998). Thus, investigating language acquisition and processing in a multilingual environment might be interesting as well mainly because this environment can fasten language development and language change. Crystal (2004) has already stressed this view stating that linguistic changes in the multilingual environment (especially in the Internet) are more rapid than at any previous stage in language history. Moreover, multilingualism is currently a widespread phenomenon in the world, since most people are potentially multilingual by nature and accordingly multilingualism is considered as the normal state of linguistic competence (Aronin and Singleton 2008; Auer and Wei 2007; Cook 1992; Grosjean 1982; 2010; Hammarberg 2010). Relatedly, Crystal (2003:14) believes 'in the fundamental value of multilingualism, as an amazing world resource which presents us with different perspectives and insights, and thus enables us to reach a more profound understanding of the nature of the human mind and spirit'.

Therefore, increasing our knowledge of the multicompetent mind and comparing it with findings of the theoretical tenets of single and dual mechanism approaches of the monolingual mind appear to be a potentially fruitful. The present study attempts to make a humble contribution to this line of investigation by exploring recent linguistic developments and movements in English verbal system within an intensely multilingual environment of the Internet. More specifically, as an additional testing ground to evaluate single and dual mechanism approaches, this study aims to investigate whether (ir)regularization processing takes place over time by running a corpus study in the Internet. Following the view of Crystal (2004), in this multilingual setting, the linguistic developments and changes are expected to take place faster than usual. It is in a way a laboratory when language development and change may take place much faster than in any monolingual communities (see chapter 3 for more details).

Another important plausible view to remember concerning the relevant roles of children and adults in diachronic changes: whether children have a vital role in diachronic changes or whether changes may take place in the adults' language as well. In the next section, more details about this issue will be provided.

2.3 Roles of children and adults in diachronic changes

In recent years, there has been a heated debate in the literature concerning the respective roles of children and adults in diachronic changes. The arguable standpoints are, on the one hand, that language acquisition by children is essential to understanding diachronic change. Most generative linguists adopt the child-based theory. They assume that child language acquisition is the locus where grammar change occurs because of a failure in transmission of certain linguistic phenomena over time. On the other hand, sociolinguists reject that this essential role of children in diachronic changes claiming that changes may take place in the language of adult speakers as well (Labov 1994; Croft 2000; Aitchison 2004). In the same respect, Weerman (1993); Hróarsdóttir (2009) and Longobardi (1999, 2001) argue that the two different viewpoints are not necessarily controversial, these two opinions can be true at the same time as they valid for different aspects of the diachronic change. In the next sections, I will provide more details about this debate.

2.3.1 Internal forces for diachronic change

For many decades, the relation between language acquisition and a historical change has fascinated historical linguists who insist that the process of language acquisition can play a central role in historical development (Müller 1890; Paul 1890; Halle 1962; Chomsky and Halle 1968; Andersen 1973; Lightfoot 1979; Chomsky 1986, 2000 among other). The most important observation for L1 acquisition is that healthy children acquire their native language perfectly in a relatively short period of time. Chomsky (2000: 122) points that the process of language acquisition is directed internally. For him, it is something that happens to a child, not that the child does (Chomsky 2000: 7). Hence, children, who are gifted with certain innate capacities, can automatically acquire knowledge of a language. Basing on this, generative linguists claim that the locus for change is inter-generational language transmission. Language change, therefore, relates to a different parameter setting by the new generation as a result of reanalysis.

Following generative grammar, two notions of language should be distinguished: I(nternal) language and E(xternal) language. I(nternal) language refers to innate grammatical system that is a part of speaker's grammatical knowledge. Hence, I-language is an internal (individual) system embodied in people's mind, whereas E(xternal) language refers to observable linguistic expressions produced by a community in communication. Thus, it is the population of utterances in a speech community (see Croft 2000: 26). E-language, on the one hand, is flexible and constantly changing in the lifetime, quite often in unsystematic ways. I-language, on the other hand, changes more systematically during childhood. The most significant changes of language history are concerned with I-language changes that happen during the critical period of L1 acquisition.

Therefore, generative linguistics claim that humans have parameters that belong to the innate language faculty and children (as the instigators of change) reconstruct their grammars based on the linguistic input they receive around them. They first base themselves on the output of their parent's grammar, and then they may construct a grammar that is different from their parents. Once the children reach the adult age, their grammars become the new basis for the next generation of children to reconstruct their grammars. So, language acquisition is the medium through which language change is transmitted over time. This iterative process that specifies the dynamics of a formal model of language change is shown in figure 9 below.

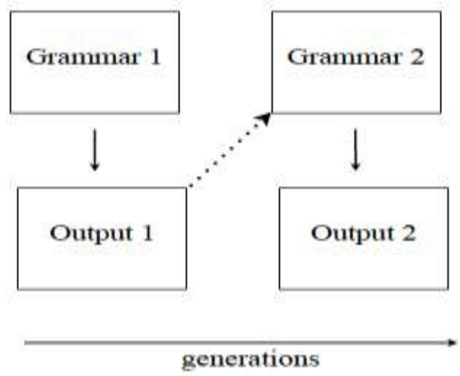


Figure 9: The model of language change (based on Andersen 1973: 767 in Postma 2017)

In this model of language change, Andersen (1973) argues that grammar 1 itself cannot be duplicated to the next generation. Instead, it is acquired via its output. The next generation acquires their grammar through exposure to the set of utterances produced by adults. In this stage, language acquisition can result in language change whenever children face imperfection in the process of learning. This happens if a child guesses a grammar that approaches grammar 1 and makes a tentative correction using an adaptive rule. This model with adaptive rules has already been suggested by Halle (1962:64 in Postma 2017) when he states:

The language of the adult – and hence also the grammar that he has internalized– need not, however, remain static: it can and does, in fact, change. I conjecture that changes in later life are restricted to the addition of a few rules in the grammar and that elimination of rules and hence a wholesale restructuring of his grammar is beyond the capabilities of the average adult (Halle 1962: 64).

According to these considerations, a model of language acquisition should be integrated within a model of language change as an essential part. Hence, when describing any diachronic changes, it is necessary to take into consideration of how those changes occur from the language-learning perspective.

Contra the view of most generative linguists who mainly account for the essential role of children to understanding diachronic change, most of sociolinguists mainly focus on E-language as a crucial force in explaining linguistic changes over time (Weinreich, Labov and Herzog 1968; Croft 2000; Aitchison 2004). They argue that adults and not children, are the drivers of diachronic changes. Hence, they focus on the external language change within social groups where innovations of linguistic variation occur in adults' languages. More details concerning this view are provided in the next section.

2.3.2 External forces for diachronic change

Sociolinguists, unlike most generative linguists, claim that children do not play a vital role in diachronic changes (Labov 1994; Willis 1998; Croft 2000; Aitchison 2004). In this respect, Aitchison (2004: 216) states that:

The belief that children initiate change was a hopeful guess made by linguists to whom the whole process of change was mysterious. In fact, similarities between child language and language change are largely illusory. Children are unlikely to initiate change, since change is spread by social groups, and babies do not have sufficient group influence to persuade other people to imitate them (Aitchison 2004: 216).

Therefore, Aitchison assumes that language change is concerned with something happens in mastering a language well after childhood, hence it is not L1 acquisition that is relevant for diachronic changes. Relatedly, Croft (2000: 57) argues that "[i]f linguistic variables are a part of the grammar…, then changes in use are changes in grammatical knowledge. In other words, changes can occur in the grammar of adults in the course of language use". Croft claims the primary locus of language change can be in the selection of linguistic expressions and utterances in social contact.

Additionally, Lev Michael (2014: 1) claims that "[l]anguage change results from the differential propagation of linguistic variants distributed among the linguistic repertoires of communicatively interacting individuals in a given community". So, most sociolinguists try to display how social relations a community may lead to spreading of certain innovation that may cause certain linguistic changes in a given language. Relatedly, Milroy and Milroy (1985) do not consider innovation as a change in itself. This innovation can only become a change when it is first adopted and then diffused by members of a community. External social factors like age, sex, prestige, ethnicity, etc. are generally expected to control the diffusion of innovations.

So far, the main concern is with the diffusion of innovations (E-language changes), but how the diffusion of I-language changes may occur? Willis (1998: 47- 48) argues that parameter settings associated with I-languages are unable to diffuse themselves, because the diffusion of I-language changes is very dissimilar to the diffusion of E-language changes. He claims that:

A parametric shift spreads in so far as the change of parameter setting in one speaker or group of speakers tilts the trigger experience of children towards the new setting. That is, once one speaker shifts to the new setting, the amount of data in favour of the old parameter setting falls, whilst the amount of data in favour of the new parameter setting rises (Willis 1998: 47- 48).

Some linguists like Weerman (1993), Longobardi (1999, 2001) and Hróarsdóttir (2009) adopt the view that both I-language and E-language changes are two necessary steps in order to have an explanatory success of a diachronic change: E-language change (innovations of variation caused by language contact or other changes in a community) and the following I-language change (a biological (internal) change in L1 acquisition). This view is well expounded in Weerman (1993). In Weerman's theory, the source of language change situates in adults who distort their language under pressure of L2 acquisition. More details about this viewpoint will be presented in the next section.

2.3.3 Both internal and external forces for diachronic change

As previously mentioned, generative linguists mainly focus on I-language changes where children are the real agents of diachronic changes. Hence, they only assert the precise nature of the parameter change disregarding the prior E-language change where innovations of variation may occur in adults 'language. However, sociolinguists mainly focus on E-language changes and its diffusion where adults are considered as the instigators of change. Various scholars (Weerman 1993; Longobardi 1999, 2001; Hróarsdóttir 2009 among others) assume that I-language changes may only take place when there are changes in language use of the previous generation. These changes are caused by certain external forces like language contact that may pave the way for a new interpretation. Consequently, they assume that to have an explanatory success of a diachronic change, both I-language changes (acquisition-based grammar changes within the language of children) and E-language changes (innovations of variation within the language of adults) must be accounted for through integrating E-language innovations into a stable I-language during L1 acquisition.

Weerman (1993) tries to prove that one crucial observation that shows the contrast between L1 and L2 acquisition is that the parameters of one's mother tongue become fixed during childhood. So, no change is possible for these parameters later on. In this vein, Halle 1962; Lightfoot 1999 and Yang 2002 assume that peripheral rules[2], introduced by adults or second language (L2) acquirers, can be constructed into I-language by a next generation. relatedly, Weerman (1993) shows that a change in terms of parameters is possible if children, in turn, base themselves on the output of the grammar of L2. Thus, children may acquire certain peripheral rules from the input of adults around them and consider these peripheral rules as a basis to change their parameters. So the children are partially the instigators of the language change, as they receive their input from the adults who cannot reset the parameters they set as children.

In figure 10, Weerman puts this discussion in systematically in which A, B, C are considered as central positive evidence for parameters that are set during childhood. These parameters are fixed, but they can be reset only when children are confronted with peripheral rules of L2 acquisition, like A, B, D. Afterwards, children can use these peripheral rules as the basis for setting their parameters. In this model, adults are considered as the central agents of E-language change, since they introduce instability and innovations of variation. Children, however, are considered as the vital agents in the acquisition-based grammar change because of their innate instincts that enable them to parse, generate and create their language according to the constraints of Universal Grammar.

[2] Chomsky (1981) distinguish between phenomena that are central to the grammar and phenomena that are peripheral. He uses the "core grammar" for the phenomena that are central to the grammar versus "peripheral rules" for the phenomena that are peripheral to the grammar.

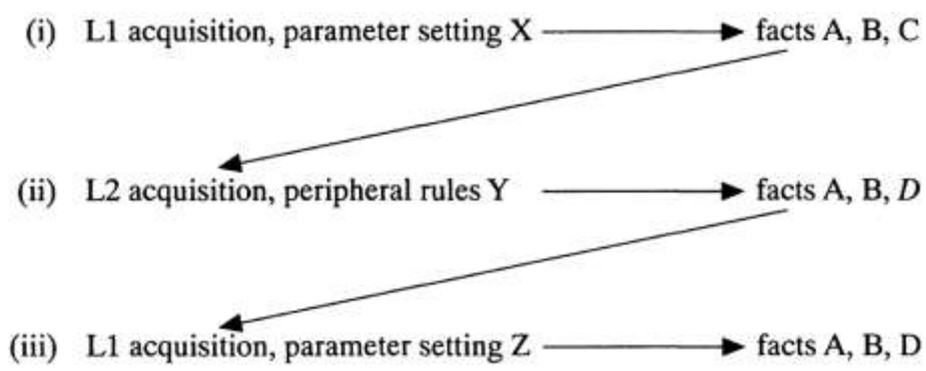


Figure 10: A model of a language change with the use of a peripheral rules solution by Weerman (1993: 910)

Weerman postulates that adults who have acquired a perfect grammar 1 may develop it by adding peripheral rules under certain circumstances such as language contact. When exposed to this kind of a grammar (grammar 1 + peripheral rules), children may re-set their parameters in order to be harmonious their parent's output and the requirements of universal grammar.

To support this view, Weerman (1993) claims that the main cause of the change from OV to VO in English is due to contact with L2 speakers. From about (800 to 1050), large numbers of Vikings settled in England in the north and east (the area where the change started). These large numbers of L2 speakers owned most of the circumstances needed for a language change: they were large enough in number, they have a good contact with the original inhabitants and their language had many parallels to Old English. As the Vikings had a language with VO word order and they faced to understand ambiguous clauses in English with OV word order, they assumed their word order in English to solve the problem. Gradually children began to reset their parameters when they were confronted with the VO input; this led to the change of English word order slowly. The same change didn't happen in Dutch, because Vikings did not colonize it; that is why Dutch word order is still OV. The unchanged word order in Dutch can present an extra evidence that there must have been an external trigger in a diachronic change.

In the context of the current study, it is worthwhile to consider a relation between the acquisition of morphological inflection and language change in a case where two (or more than two) languages are in contact with one another. In such cases, it appears that adult learners are less successful in mastering inflection than young learners (Johnson and Newport 1991; Hyltenstam and Abrahamsson 2003; Blom et al. 2007 among many others). This phenomenon may have an effect on the result of the acquisition process of the learners of the next generation. Because adult learners lack the ability to acquire the target system fully, their language will differ at some points from the previous system and this variation will cause differences in the input for the next generation. Hence, a less consistent input may denote less evidence for specific inflectional contrasts and this may cause a possible loss of such contrasts altogether. According to this observation, it is predicted that there is a correlation between the number of adult learners in a language community and deflection[3]. Some linguists (Trudgill 1986; Kroch and Taylor 1997 among others) claim that the degree of language contact correlates with the degree of deflection; more language or dialect contact implies more deflection. Relatedly, in bilingual and multilingual situations, Trudgill (1989: 228-9) claims that people may try to loose “marked or complex variants” in favor of “unmarked, or simpler forms”. This fact has previously been noticed by Jakobson (1929). The role of people in these situations has also been underlined Thomason (2003: 692) who observes that people “fail to learn some features of the [target language], usually features that are hard to learn for reasons of universal markedness”. This observation is in equivalent with Trudgill’s remarks on simplification.

Interestingly, in the multilingual internet, people inflect the words differently using various types of unmarked and marked forms (Crystal 2004). For instance, some people inflect English irregular verbs with unmarked (regular) forms e.g., *speak-speaked* and *spend-spended.* In addition, some new words are inflected with regular forms e.g., *google-googled, upload-uploaded*. However, marked (irregular) forms are used as well. For instance, the irregular verbs *sprung, brang* and *stunk* are used as past forms of *spring, bring* and *stink* respectively. Also, some new words are inflected irregularly e.g., *twat* or *twot* is used as a past form of the verbs. Similarly, new nouns e.g., *vax* and *bix* are pluralized as *vaxen* and *bixen* respectively (Crystal 2011: 67). Sometimes, inflected expressions of certain words are a mixture of both regular and irregular forms

[3] Deflection refers to a property of the Germanic languages that show a tendency to reduce or get rid of their inflectional contrasts (Weerman 1989; Aalberse 2009).

e.g., *matrix-matrixes* or *matrixen*. As previously mentioned, in the internet environment, linguistic developments and changes are expected to at a much faster speed than ever before (Crystal 2004).

In the current study, I will make a contribution to the single-dual debate by investigating language use in Contemporary English. To this end, by running a corpus study based on data from the internet environment, this study aims at investigating to what extent current uses of regular and irregular verbs may provide evidence for single or dual mechanism by focusing on the question whether only regularization processes apply (favoring a dual mechanism approach) or whether processes of irregularization can also be attested (favoring a single mechanism approach). In the next chapter, the methodology that will be used to carry out this study will be displayed.

3 Chapter Three: Methodology

In this chapter, I will illustrate the methodology that I will use to explore English (ir)regularization processes on the synchronic and diachronic levels by running a corpus study. It consists of five main sections. Section 3.2 provides the necessary review of the theoretical background for analyzing the data of the present study. This is followed by the presentation of the research questions and the main related predictions of single-dual mechanism approaches in section 3.3. In sections 3.4 and 3.5, I will sketch out the motivation of choosing the internet environment and WebCorp as a Linguistic Corpus for the present study. Finally, in sections 3.6, I will describe the way of selecting the data of this study and the procedures that will be followed in analyzing this data.

3.1 Summary review of the theoretical background

As previously mentioned in the theoretical background chapter, the fundamental issue behind the dispute of language processing between advocates of single-dual mechanism approaches is whether the human language system exploits two cognitive mechanisms (rules and storage) or a single mechanism (either rules or storage) (Chomsky and Halle 1968; Rumelhart and McClelland 1986; Pinker 1999 among others). From the dual mechanist perspective, IVs are fully stored in the associative memory as memorized pairs of words, while RVs are generated by rules (the addition of the *-ed* suffix to a verbal stem). Consequently, it is predicted that IVs are sensitive to the properties of full form storage in the associative memory and hence exhibit frequency effects and phonological similarity effects. However, RVs processed via the rule system, are not expected to be sensitive to these effects. Hence, different mechanisms must underlie the production of regularization and irregularization processes. These two processes should be at different rates: the regularization rate should be higher than the irregularization rate. IVs with low frequency are expected to be regularized more often than IVs with high frequency at an average rate of 4.2% (Marcus et al. 1992). Yet, the irregularization rate is expected to be rare, only 0.2% in the study of Xu and Pinker 1995.

The higher rate of regularization compared to irregularization in the dual mechanism view is in conflict with predictions by single-system approaches. From the single mechanism viewpoints, it is predicted that verbal changes are bidirectional, towards both regularization and irregularization. This is due to the hypothesis that both RVs and IVs are processed via a single mechanism: either rules or storage. Therefore, they propose that similar mechanisms underlie the production of regularization and irregularization, hence both processes are expected be at similar rates. From connectionist perspectives, RVs as well as IVs are stored in the associative memory. Thus, both RVs and IVs are expected to exhibit the impacts of frequency and phonological similarity. Regularization and irregularization should therefore display the impacts of these effects. Followers of rule-based models, however, hypothesize that both RVs and IVs are handled by rules. They are expected to be frequency-insensitive. Hence, regularization and irregularization must exhibit display frequency-insensitivity.

Empirical and theoretical studies surrounding the single-dual mechanism debate have mainly focused on the acquisition of the English past and perfect forms (Rumelhart and McClelland 1986; Pinker and Ullman 2002; Yang 2002; Lieberman et al. 2007, Michel, et al. 2011 among other). This is because English past and perfect formations appear to comprise two distinct systems: regular (add -*ed* to a verbal stem e.g., *play-played-had played*) and irregular (the past and perfect forms are produced apparently in idiosyncratic ways e.g., *write-wrote-written, mean-meant-meant* or *cut-cut-cut*). RVs and IVs appear to behave separately from each other, as they are influenced by the syntactic, semantic or phonological properties of English (Pinker 1991). Accordingly, regular and irregular past and perfect forms, as well as instances of (ir)regularization, can formulate a good ground for investigating whether two distinct mechanisms are at work (as predicted by the dual mechanism approach) or whether a unary mechanism must underlie the production of these inflected forms (as proposed by single mechanism approaches).

Up to date, empirical studies of language processing have not provided fully findings that clearly confirm the predictions of either approach. This study is a contribution to this debate; its general aim is to investigate to what extent current uses of RVs and IVs may provide evidence for/against single-dual mechanisms. It mainly focuses on the question whether, on synchronic and diachronic levels, only regularization processes apply (favoring the dual mechanism approach) or whether processes of irregularization can also be attested (favoring single mechanism approaches). To this end, more detailed research questions are formulated in the next section.

3.2 Research questions

With the focus on language acquisition and language change, I aim to investigate current movements in English verb system synchronically and diachronically. I aim to check whether verbal changes are unidirectional (towards regularization) or bidirectional (towards both regularization and irregularization). To do so, I run a corpus study to explore whether there is a relationship between:

- (ir)regularity and word frequency on a synchronic level
- (ir)regularization and word frequency on a synchronic level
- (ir)regularization and word frequency on a diachronic level

If a relationship between irregularity (or regularization) and word frequency on synchronic and diachronic levels will be attested in this study, this (in principle) would provide evidence for the dual mechanist view arguing against single mechanist views. For this purpose, the following questions are addressed:

A- On the synchronic level

1. Are IVs generally more frequent than RVs in the past and perfect forms in Contemporary English?

The dual mechanism approach predicts a relationship between irregularity and word frequency, as IVs are fully stored in the associative memory and hence they should be frequency-sensitive. Whereas RVs are generated by a rule and they have to display frequency effects. Single mechanism approaches, however, predict that there is no clear relationship between (ir)regularity and word frequency, as there is no principled difference between RVs and IVs. In this investigation, I aim to check whether the relationship between (ir)regularity and word frequency displays symmetries (favoring single mechanism approaches) or asymmetries (favoring the dual mechanism approach).

2. Do regularization processes take place in Contemporary English? If so, are IVs with low frequency regularized more often than IVs with high frequency in the past and perfect forms?

3. Do irregularization processes take place in Contemporary English? If so, are IVs with low frequency regularized more often than IVs with high frequency in the past and perfect forms?

Proponents of the dual mechanism approach predict that there is link between word frequency and regularization processes. Accordingly, IVs with low frequency are supposed to be regularized more often than IVs with high frequency because of the weaker memory traces of IVs with low frequency possess and thus their resistance to regularization processes is less than the one with high frequency. Nevertheless, followers of single mechanism approaches predict that there is no link between word frequency and regularization processes due to their hypotheses that both RVs and IVs are processed by a single mental mechanism (rules or storage). Therefore, the same mechanism must underlie the production of both regularization and irregularization processes and hence their rates expected be alike. Furthermore, instances of irregularization can display phonological neighbourhood effects as predicted by dual and connectionist models. This is due to their hypothesized full-form storage of IVs in the associative memory. Yet, from connectionist perspectives, irregularization rates can be in a range from 3.2% to 23.5% (Rumelhart and McClelland 1986; Plunkett and Marchman 1991; Sproat 1992; Plunkett and Marchman 1993). This range is much higher than the irregularization rate presented in the empirical studies of dual mechanism models; like only 0.2% in the study of Xu and Pinker (1995). On the synchronic level, I aim to explore whether or not there is a relationship between (ir)regularization processes and word frequency in Contemporary English to generate evidence in favour with or against single and dual mechanism approaches.

Furthermore, I explore whether there is a link between word frequency and regularization processes in a limited set of doublet verbs that can be conjugated as both IVs and RFs in the English language. For instance, verbs like *burn* and *dream* are conjugated as *burned*/*burnt* and *dreamed*/*dreamt* in the past and perfect forms. Some of these verbs have no vowel change between IVs and RFs like, *learned*/*learnt* and *spoiled*/*spoilt*, others undergo vowel change as in *leaped*/*leapt* and *lighted*/*lit*. Here, I aim to investigate the impact of vowel change on the retention of regularization processes. I explore whether there is a relation between vowel change and regularization processes by checking to what extent vowel change (or the absence of it) may possibly account for the variation in such verbs. Intuitively, I predict that verbs with no vowel change will meet with less resistance in these processes (and consequently display more RFs) than verbs with vowel change. The fewer differences between IVs and RFs there are, the less resistance for regularization language users have, and hence the more RFs there are. Thus, the following question is addressed:

4. Do regularization processes occur more frequently in the cases where IVs and their corresponding irregular forms (e.g., *learn-learnt/learned*) show no vowel change in Contemporary English?

De Clerck and Vanopstal (2015) have already examined the impact of vowel change on the retention of regularization processes in 11 doublet verbs namely *burn, dream, dwell, kneel, lean, leap, learn, smell, spell, spill and spoil* (For more details see section 2.3.3). The analysis of this study reveals that there is a relationship between the salience of vowel change and regularization processes: verbs with no vowel change are regularized more often than verbs with vowel change. I aim to explore whether this trend can still be attested in the selected sample of this study.

B. On the diachronic level

5. Are verbal changes towards regularization taking place constantly over time in Contemporary English?
6. Are verbal changes towards irregularization taking place constantly over time in Contemporary English?

As has been stated before, language change is attested in both directions: unidirectional in which IVs commonly evolve to become more general (Fries 1940; Pinker 1999; Lieberman et al. 2007; Michel et al. 2011 among other), and bidirectional in which language changes are in the direction of both regularization and irregularization (Peters 2004; Fretig 2009; Cuskley et al. 2014 among others). So, our aim is to investigate diachronically whether verbal change is unidirectional (only towards regularization) or bidirectional (towards both regularization and irregularlization). If verbal changes are unidirectional, this will be in favour of dual mechanism approaches, as only IVs are regularized in case they are not heard often enough to be stored in the associative memory. If verbal changes are bidirectional, this will be in favour of single mechanism approaches. In rule-based models, regularization is not in principle different from irregularization; it is a matter of a change from one rule to another. One rule is called the regular rule which controls the majority of verb types. Whereas, a collection of minority rules is applied to IVs. Nevertheless, followers of connectionist models assume full-form storage of both RVs and IVs in the associative memory. Consequently, they hypothesize that same mechanisms must underlie the production of both regularization and irregularization processes and fittingly their rates are predicted be similar (For more details see chapter 2).

In the current study, I assume that verbal changes of (ir)regularization are taking place more rapidly than usual in the multilingual environment; particularly in the internet space. My hypothesis is based on Crystal's (2004) claim concerning language changes in the multilingual internet that appear to be taking place more quickly than at any earlier time in history of linguistics. In the next section, I will provide more details about multilingualism especially in the internet environment.

3.3 Multilingualism in the internet

Multilingualism refers to the phenomenon of using "three or more languages, either separately or in various degrees of codemixing" (McArthur 1992: 673). Wei (2008:4) claims that "anyone who can communicate in more than one language, be it active (through speaking and writing) or passive (through listening and reading)" to be multilingual. Therefore, multilingual individuals can easily switch between languages (Lüdi and Py 2009) or use two or more languages on a daily basis (Grosjean, 2010). Nowadays, multilingualism is a widespread phenomenon in the world, as the majority of people are considered as multilingual and not monolingual (Aronin and Singleton 2008; Auer and Wei 2007; Cook 1992; Grosjean 1982, 2010). Relatedly, Hammarberg (2010: 92) argues that "humans are potentially multilingual by nature, and that multilingualism is the normal state of linguistic competence". Therefore, it is essential to explore current linguistic phenomena from a multilingual viewpoint.

Due to the expansion of British colonial power in the 19th century and the rise of United States as the leading economic superpower of the 20th century, English has been adopted as the preferred language for international communication by 22 multilingual countries that account for about 5000 of the world's 6000 languages (Crystal 1997, 2003). In the same respect, Graddol (1997) displays the estimated population of three types of English speakers: 375 million L1 speakers, 375 million L2 speakers and 750 million foreign language speakers. The spread of English has already been exemplified in three circles by Kachru (1985: 5) (see figure 11). This model has been generally considered as a useful approach, though some countries do not precisely fit into it. These three circles are categorized as follows:

- the inner circle involves the countries that are traditionally regarded as the bases of English, where English is L1 for the majority of the populations: UK, USA, Ireland, Canada, New Zealand, Australia. However, English is not the only language spoken in these countries, as it is in contact with other languages. This is the result of immigration in these countries.
- the outer circle contains those countries where English is not L1 of the majority of the population. Instead English is considered as L2 that is used at the institutional level as the result of colonization. It comprises Singapore, India, Malawi and other territories
- the expanding circle includes those countries where English has no official status and is taught as a foreign language or as an international language, although they did not colonize by members of the inner circle. It involves China, Japan, Greece, Poland. In addition, as the name of this circle suggests, other states may join this circle in the contemporary scene.

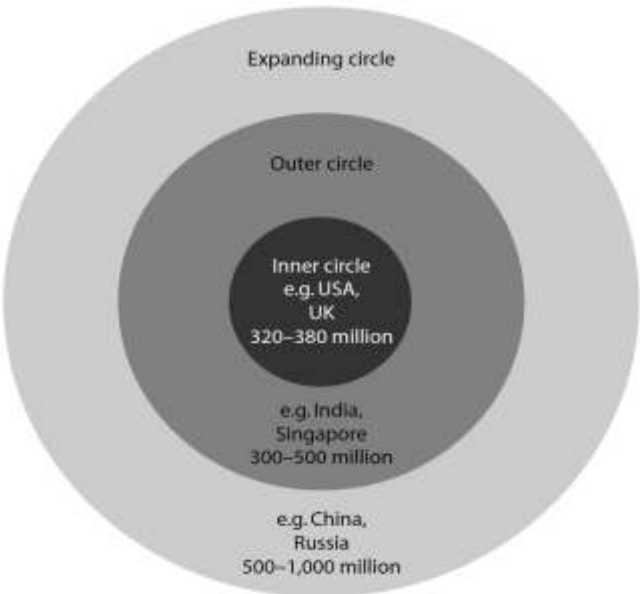


Figure 11: The three ‘circles’ of English (based on Kachru 1985: 5)

Among these three circles, the expanding circle is the fastest growing circle due to technological development and interesting international contact (Gnutzman 2000). In this expanding circle, English is used as a lingua franca that refers a language that is adopted as a common language among speakers whose native languages are different (Crystal 2003; Graddol 2006). However, this does not mean to exclude outer or inner circle speakers from using as English a lingua franca. Instead, in certain circumstances and places, speakers of the three circles need to make adjustments to their local English variety for the benefit of their interlocutors when they take part in lingua franca English communication (Danet and Herring 2007). It is a common view that the internet is one of a notable place for diversity and creativity where many people from different linguistic backgrounds are able to be in contact from distance. Those people can quickly learn to adapt their language to meet the demands of the new situations and to exploit the potential of the new medium so creatively to form new areas of expression. Additionally, they are able to use different strategies to guarantee successful virtual communication when they find themselves interacting with each other specifically with the use of social networking services like Blog, Facebook and Twitter. These social networking services are considered as the most productive environments of English word formation (Herring 2007; Evans 2009; Diemer 2011).

Despite the pervading nature of multilingualism in the last few decades, the focus of most empirical studies in different fields of linguistics are mainly on how human language works in monolingual and bilingual environments (Cenoz et al., 2003). As mentioned previously in the theoretical framework, language contact is considered as a crucial factor that plays a central role in language change. I predict that the Internet (as intensely multilingual medium of language contact) is a suitable place where language change may take place more rapidly than usual. It is in a way a laboratory when language change may take place much faster than in any monolingual communities. Accordingly, the present study aims to contribute to this line of investigation by running a corpus study extracted from the environment of the internet to explore recent linguistic movements in English verbal system.

3.4 WebCorp as a linguistic corpus

To investigate the direction of current movements in verbal system synchronically and diachronically, the World Wide Web (WWW) is used as a linguistic corpus for several reasons. First, as has been stated before, this linguistic environment can be considered as a multilingual medium of language contact where language change takes places more rapidly than elsewhere. Second, it can be the only source for examples of very rare usages and linguistic constructions (Mair 2010). Third, it is a self-updating monitor corpus and hence an excellent source for recent grammatical innovations. It can provide data on a scale that is simply not matched by corpora such as BNC (British National Corpus)[4] and COCA (Corpus of Contemporary American English)[5] (Keller and Lapata 2003, Diemer 2011). Several years ago, Kilgarriff (2001) has first drawn our attention to the value of the web as a linguistic resource in different fields such as sociolinguistics and natural language processing:

The corpus resource for the 1990s was the BNC. Conceived in the 80s, completed in the mid-90s, it was hugely innovative and opened up myriad new research avenues for comparing different text types, sociolinguistics, empirical NLP, language teaching and lexicography. But now the web is with us, giving access to colossal quantities of text, of any number of varieties, at the click of a button, for free. While the BNC and other fixed corpora remain of huge value, it is the web that presents the most provocative questions about the nature of language (Kilgarriff 2001:344).

Nevertheless, there is currently a debate about using the web as a corpus. Some of the main criticisms of using the web as a corpus are:

1. The web is “dirty” with numerous erroneous forms (Kilgarriff and Grefenstette 2003: 342).

[4] BNC is a 100 million-word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of British English, both spoken and written, from the late twentieth century.

[5] COCA is the freely searchable 450-million-word corpus, which is considered the largest corpus of American English currently available. It contains a wide array of texts from a number of genres.

2. The large amount of duplication on the web can distort counts of the number of hits (Lüdeling et al. 2007; Fletcher 2007).
3. Search engines may be unreliable. For example, one may find certain returning hits that are not actually on the page itself, but rather contained in a link to the page (Keller and Lapata 2003).
4. Results cannot be arranged in the format of easily readable concordance lines.

In spite of these criticisms, I will use the WWW as a corpus for my current study for three reasons. First, some of the "dirty" forms referred to by Kilgarriff and Grefenstette (2003: 342) can be significant evidence of language change and hence are potentially forms of some linguistic interest. Second, as the aim of the present study is to investigate the developments and changes of the verb system in Contemporary English. I assume that verbal changes may occur more than usual in the multilingual environment of the WWW (see section 3.4 for more details). In this respect, Crystal (2004: 92) claims that the rate of language change appears to be occurring across this environment 'much faster than at any previous time in linguistic history' (see section 3.4 for more details).

Third, to minimize the risks of using the web as a linguistic corpus, the WebCorp project, based at Birmingham City University is chosen for this study. This project is created by the Research and Development Unit for English Studies (RDUES) to facilitate specific use of the WWW as a linguistic corpus. WebCorp contains the WebCorp Linguist's Search Engine (WebCorp LSE) from which the data of the current study is extracted. WebCorp LSE refers to a new tailored linguistic search engine that is crawling and processing the WWW to build 10-billion-word text corpora (Kehoe and Gee 2007). Different types of corpora are found in this search engine such as synchronic English Web Corpus and diachronic English Web Corpus. Certain linguistic tools are available to help the users making more restriction for the corpus such as case insensitive, sentence position, word filter, wildcards and part-of-speech (POS). Additionally, results are shown as concordance lines in KWIC (key word in context) format. And finally, it is also allowed to enhance the sentence boundary detection, date identification, 'junk' removal, and other statistical analysis options (Kehoe and Gee 2007).

With the tools of WebCorp, if has become possible to attempt a trawl of the WWW effectively because of the insufficiency of evidence in existing corpora for rarer or newer linguistic items and features (Bergh et al. 1998; Brekka 1999, 2000).

3.5 Data selection and procedures

In order to investigate the debate of morphological processing (storage versus composition) in the current English verb system, the present corpus-based study makes use of the WebCorp corpus. Owning to its design, this corpus is most suitable for our purpose of examining ongoing verbal changes in Contemporary English, given that it contains synchronic and diachronic data needed for this investigation (see section 3.5 for more details). The selection of the verb sample and the procedures of analyzing this sample to answer the research questions in the present study (see section 3.3) will be shown in the following case studies:

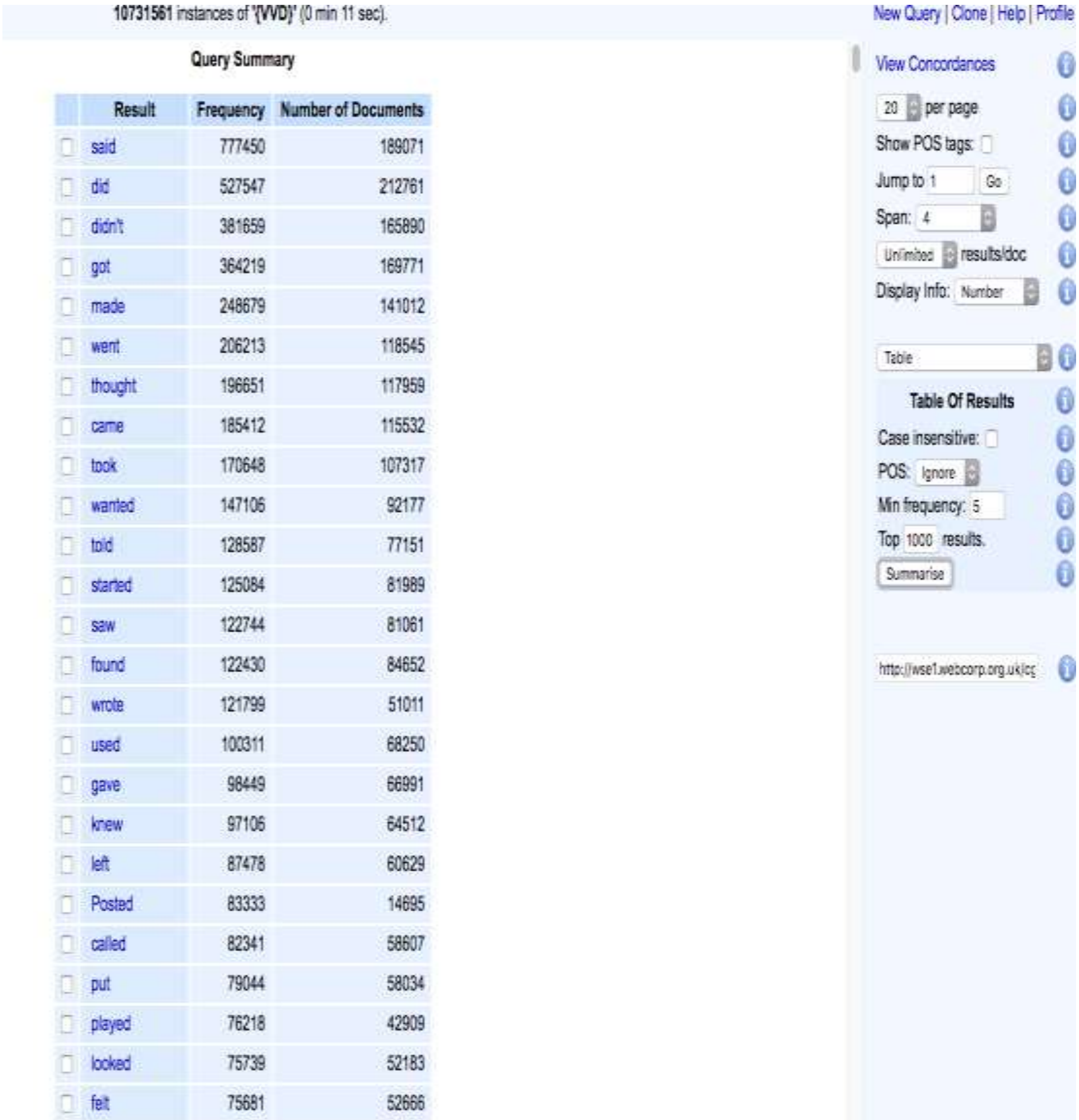
3.5.1 The synchronic snapshot

Here, in three case studies, I address the way of selecting the verb sample and the procedures that will be followed to explore word frequency effects on (ir)regularity, regularization and irregularization in the selected sample. Furthermore, the data selection and the procedures to explore the relationship between regularization and the salience of vowel change are presented in the case study of section 3.6.1.3.

3.5.1.1 Case 1: (ir)regularity and word frequency

I aim to detect and compare frequency effects for RVs and IVs. Frequency effects are taken as evidence for storage (see chapter 2 and section 3.2 for more details). If frequency effects are observed for IVs but not RVs, this may suggest that IVs and RVs are processed by different mechanisms (storage for IVs and rules for RVs). However, if frequency effects are detected (or not) for both IVs and RVs, this may indicate that both RVs and IVs are processed by similar mechanisms (either storage or rules). To explore these assumptions, I make a comparison between verbs with highest word frequency and verbs with lowest word frequency selected from the corpus of this study (WebCorp) following these steps:

1. A search for verbs in a simple past tense is made to reach all possible RVs and IVs in WebCorp. In this search, I insert the part-of-speech tag to reach these verbs in a past simple tense with the selection of {VVD}in WebCorp. The reason behind choosing a simple past tense for this search is that a verb form can be regular or irregular in this tense (like *I walked/I slept*), not in a simple present tense (like *I walk/ I sleep)*.
2. From this search, I obtain a list of 10,731,561 instances in a past tense. The first 25 verbs in this list are shown in figure 12 below.



Result	Frequency	Number of Documents
said	777450	189071
did	527547	212761
didn't	381659	165890
got	364219	169771
made	248679	141012
went	206213	118545
thought	196651	117959
came	185412	115532
took	170648	107317
wanted	147106	92177
told	128587	77151
started	125084	81989
saw	122744	81061
found	122430	84652
wrote	121799	51011
used	100311	68250
gave	98449	66991
knew	97106	64512
left	87478	60629
Posted	83333	14695
called	82341	58607
put	79044	58034
played	76218	42909
looked	75739	52183
felt	75681	52666

Figure 12: The list of 10,731,561 instances in the past tense from WebCorp

3. From this list, the top 10,000 verbs are extracted in which a minimal word frequency[6] of a verb is not lower than 2.
4. The 10,000 list contains some of unwanted results. This list is filtered (cleaned) manually to remove all unwanted hits before selecting the verb sample of the study. For example, the verb *do* has two forms in the past tense: *did* and *didn't* (see figure 12). The one with negation (*didn't)* is omitted from the list.
5. Then, I select the top 125 RVs and the top 125 IVs versus the bottom 125 RVs and the bottom 125 IVs from the filtered list [7]. In total, 500 verbs (250 RVs versus 250 IVs) are gathered (see table 7 below and appendix 1).

Table 7: The verb sample of the study: the top 125 RVs and the top 125 IVs versus the bottom 125 RVs and the bottom 125 IVs from the 10,000 list

<table>
<tr><th>Type</th><th>Verbs with high frequency</th><th>Verbs with low frequency</th></tr>
<tr><td>RVs</td><td>125</td><td>125</td></tr>
<tr><td>IVs</td><td>125</td><td>125</td></tr>
<tr><td rowspan="2">Total</td><td>250</td><td>250</td></tr>
<tr><td colspan="2">500</td></tr>
</table>

6. For each selected verb, word frequencies in the past form (like played and spoke) and the related perfect form (like played and spoken) are collected from the corpus (see appendices 2 and 3). Hence, in total, the verb sample of this investigation is 1000 verbs in both forms (500 RVs versus 500 IVs) (see table 8 below).

Table 8: The sample of 1000 verbs (RVs and IVs in the past and perfect forms)

[6] Word frequency refers to the number of occurrences of a word (or words) in a given corpus, also called frequency of occurrence(s) or token frequency.

[7] The reason behind choosing the highest versus lowest verbs is to test word frequency hypothesis stating that high word frequency of IVs is significant for their survival, as a refection of storage (Pinker and Prince 1988; Ullman 1999; Pinker and Ullman 2002 and Michel et al. 2011 among others).

Type	Verbs with high frequency		Verbs with low frequency	
	Past	Perfect	Past	Perfect
RVs	125	125	125	125
IVs	125	125	125	125
Total	500		500	
	1000			

7. Totals of word frequencies, mean frequencies[8] and relative frequencies[9] of the selected verbs split by frequency (high versus low), form (past versus perfect) and type (regular versus irregular) are calculated and displayed in tables and different types of graphs for comparative and descriptive purposes. Then, statistical models will be conducted to test the significance of the difference in frequency effects between RVs versus IVs. Finally, the obtained results will be compared to the predictions of single-dual mechanism models to determine which best fits the data and hence a conclusion will be drawn.

3.5.1.2 Case 2: regularization and word frequency

The relationship between regularization processes and word frequency will be examined in the question 2 of this study: *Do regularization processes take place in Contemporary English? If so, are IVs with low frequency regularized more often than IVs with high frequency in the past and perfect forms?* The dual mechanism approach hypothesizes that IVs with low frequency (as a refection of weaker memory traces) are regularized more often than IVs with high frequency (see chapter 2 and section 3.2 for more details). Hence to test this hypothesis, in this case study, I focus on the comparison of frequency effects for regularization instances of IVs with low and high frequencies.

[8] Mean frequency (x) is the total number of word frequencies of certain verbs (x1+x2+x3….xn) divided by the number of these word frequencies (n). $\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$. For example, mean frequency of regulars is calculated by 4,978,884 divided by 500, that equals 9,957.8.

[9] Relative frequency refers to the number of occurrences of certain verbs (like RVs) divided by the total number of occurrences of all verbs (RVs plus IVs) in the selected sample, then the product is multiplied by 100. For example, relative frequency of RVs is 4,978,884/19,762,462 = 0.26 * 100= 25.2 %.

For the purpose of this investigation, I first select the same sample of IVs with their word frequencies in both forms used in the question 1 (see appendix 2). Second, word frequencies of RFs (refer to forms obtained from adding the regular suffix *–ed* to the stems of the selected IVs e.g., *sayed*, *maked*, and *thrusted*) in the past and perfect forms are collected from the corpus (see appendix 4). Then, I compute totals of word frequencies and relative frequencies for IVs and RFs and their frequency distributions are displayed in tables and graphs of boxplots for comparative and descriptive purposes. Afterward, I will conduct a statistical model to test the significance of the difference in frequency distributions between RFs with low frequency versus RFs with high frequency. Lastly, the results of this investigation will be compared with the predictions of the models for morphological processing to determine which can best fit the data.

3.5.1.3 Case 3: the salience of vowel change and regularization

To develop a richer understanding into nature of regularization processes, I investigate the impact of vowel change on the retention of regularization processes in doublet verbs that can be both regular and irregular in the English language like *burned/burnt* and *lighted/lit* (see chapter 2 and section 3.2 for more details). To this end, 42 doublet verbs were selected from a list of 616 English IVs10. In table 9, I list the 42 doublet verbs: 21 of them show no vowel change between IVs (like *burned/burnt*) and the other 21 undergo vowel change (like *lighted/lit*). Then, word frequencies for each of these 42 doublet verbs split by type (RFs e.g., *builded, lighted* versus IVs e.g., *built, lit*) will be gathered to detect effects of vowel change for them. Totals of word frequencies and relative frequencies for these verbs are computed and displayed in tables and graphs for comparative and descriptive purposes. Finally, the significance of the difference in data distribution will be checked by conducting a statistical model.

Table 9: The selected 42 doublet verbs from a list of 616 English IVs

	Without Vowel Change	**Vowel Change**
1.	*bend-bent-bent*	*abide-abode-abidden*

[10] This is a comprehensive list of 616 English IVs, including their base form, past simple, perfect and definitions from UsingEnglish.com.

2.	*bless-blest-blest*	*alight-alit-alit*
3.	*burn-burnt-burnt*	*beseech-besought-besought*
4.	*bust-bust-bust*	*chide-chid-chid*
5.	*clap-clapt-clapt*	*clothe-clad-clad*
6.	*dwell-dwelt-dwelt*	*creep-crept-crept*
7.	*forecast-forecast-forecast*	*dream-dreamt-dreamt*
8.	*geld-gelt-gelt*	*grind-ground- ground*
9.	*gird-girt-girt*	*hang-hung-hung*
10.	*ken-kent-kent*	*heave-hove-hove*
11.	*learn-learnt-learnt*	*kneel-knelt- knelt*
12.	*pen-pent-pent*	*leap-leapt-leapt*
13.	*rid-rid-rid*	*light-lit-lit*
14.	*smell-smelt-smelt*	*shine-shone- shone*
15.	*spell-spelt-spelt*	*slink-slunk-slunk*
16.	*spill-spilt-spilt*	*sneak-snuck-snuck*
17.	*spoil-spoilt-spoilt*	*speed-sped-sped*
18.	*strip-stript-stript*	*stave-stove-stove*
19.	*sweat-sweat-sweat*	*strive-strove-striven*
20.	*wed-wed-wed*	*tread-trod-trodden*
21.	*wet-wet-wet*	*weave-wove-woven*

3.5.1.4 Case 4: irrregularization and word frequency

Here, I aim at exploring the relationship between irregularization processes and word frequency. The dual mechanism approach posits that verbal changes are unidirectional (towards regularization only) while single mechanism approaches claim that these changes occur bi-directionally (towards both regularization and irregularization) (see chapter 2 and section 3.2 for more details). In this case study, to examine these hypotheses, I will detect and compare frequency effects for irregularization instances of IVs with low and frequencies. To do so, the following steps are taken:

1. I select the same sample of IVs with their word frequencies in both forms used in the question 1 excluding suppletives[11] (*be, have, do, redo, outdo, undo, go, forgo*) (see appendix 5).
2. Word frequencies of IFs in both forms are gathered from the selected sample (see appendices 6 and 7). IFs refer to irregular forms obtained from applying certain classes of IVs to IVs of other classes e.g., *bring-brang-brung* in reference to *sing-sang-sung* (more details about these classes and the way of application will be added later).
3. I compute in tables totals of word frequencies and relative frequencies for IVs and IFs and exhibit them in tables and graphs for comparison.
4. A statistical model will be conducted to check the significance of the difference in frequency distributions between IFs with low frequency versus IFs with high frequency.
5. On the basis of the obtained results, a conclusion will be drawn.

To explore to what extent certain classes of IVs are applied to IVs of other classes, I look at IVs (without suppletives) and on the basis of vowel identity of their present, past and perfect forms, I come up with 3 different paradigms:

- One-vowel paradigm; (present=past=perfect e.g., *cut /kʌt/-cut/kʌt/-cut/kʌt/)*
- Two-vowel paradigm (present≠past=perfect e.g., *meet /mi:t/-met/met/-met/met/)*
- Three-vowel paradigm (present≠past≠perfect e.g., *ring /riŋ/-rang /ræŋ/-rung /rʌŋ/*)

Then, these paradigms are subdivided into 35 classes according to further phonological distinctions (see table 10). For example, 1C-1 refers to the class of IVs forms (e.g., *learn /lə:n/-learnt /lə:nt/-learnt /lə:nt/*) with one-vowel paradigm, but with the addition of *–t* in the past and perfect forms. Whereas, 2D-1 refers to two-vowel paradigm (e.g., *wear /weə/-wore /wɔ:/- worn /wɔ:n/*) but with the addition of *–n* in the perfect form.

---

[11] The suppletives have to be excluded, as I have to test to what extent certain classes of IVs are applied to IVs of other classes and it is hard to classify these suppletives (with little or no correlation between their past and perfect forms, e.g., *be, have, do* and *go*) into any class.

Table 10: Classes of IVs on the basis of vowel identity (35 Classes)

One-vowel paradigm			
1A - present = past = perfect			
Class	N.	Class	IVs
1A-1	1	-t → □-t -t	*put, hit, set, let, cut, bet, shut, cost, quit, cast, hurt, split, shit, broadcast, burst, spit, upset, thrust, forecast, sweat, slit, recast, bust, inset, knit, miscast, recut, typeset, intercut, typecast, wet, uppercut, podcast, offset, undercut*
		-d → □-d -d	*spread, shed, bid, rid, wed*
1B - present = past = perfect +n			
1B-1	2	-t → □-t -t+ en	*beat*
1C - past = perfect			
1C-1	3	-n → □-nt -nt	*learn, ken, burn, pen*
1C-2	4	-l → □-lt -lt	*dwell, spoil, spill, smell, spell*
1C-3	5	-p → -pt -pt	*strip, clap*
1C-4	6	-ss → -st -st	*bless*
1C-5	7	-k → □-d -d	*make, remake, unmake*
1C-6	8	-d → □-t -t	*spend, send, build, rebuild, lend, bend, overspend, gird*
1C-7	9	eɪ → eɪ +d eɪ +d	*pay, lay, overpay, repay, waylay, inlay*
Two-vowel paradigm			
2A - present = perfect			
2A -1	10	ʌ → eɪ ʌ	*come, become, overcome*
2A -2	11	ʌ → æ ʌ	*run, outrun, overrun, rerun*
	2B - present = perfect +n		

2B -1	12	eɪ → ʊ eɪ +n	*take, shake, partake, retake, forsake, betake, mistake, overtake, undertake*
2B -2	13	ɪ → eɪ ɪ +n	*give, forgive, forbid*
2B -3	14	əʊ → uː əʊ: +n	*know, throw, grow, blow, overthrow, outgrow, underthrow, foreknow, overblow*
2B -4	15	ɔː → uː ɔ: +n	*draw, withdraw, redraw ,overdraw*
2B -5	16	iː → ɔː iː +n	*see, oversee, foresee*
2B -6	17	iː → æ iː +n	*eat*
2B -7	18	ɒ → e ɒ +n	*fall, befall*
2C - past = perfect			
2C-1	19	ɪ → ʌ ʌ	*win, stick, dig, swing, spin, string, fling, cling, sling, slink, wring, unstick*
		æ → ʌ ʌ	*hang, overhang*
		aɪ +ke → ʌ +ck ʌ +ck	*strike*
		i:+k → ʌ +ck ʌ+ck	*sneak*
2C-2	20	ɪ → æ æ	*sit, babysit*
2C-3	21	æ +nd → ʊ+d ʊ+d	*stand, understand, withstand*
2C-4	22	iə → ə: +d ə:+d	*hear, mishear*
2C-5	23	aɪ → aʊ aʊ	*find, grind, unbind, rewind, unwind*
		ɪ → aʊ aʊ	*wind*
2C-6	24	iː → e e	*bleed, read, feed, meet, lead, speed, mislead, plead, breed, inbreed, overfeed*
		iː → e +t e +t	*keep, dream, leap, deal, feel, weep, mean, sleep, sweep, creep, oversleep, kneel, lean*
		ei → e +d e +d	*say, unsay*
		iː → e +d e +d	*flee*

		i: +ve → e +ft e +ft	*leave*
		əʊ → e e	*hold, bereave, uphold, withhold, behold*
2C-7	25	u: →ɒ (+d) ɒ (+d)	*shoot, overshoot, outshoot,*
		u:□□→ɒ +t ɒ +t	*lose, shoe*
2C-8	26	aɪ → əʊ əʊ	*shine, abide, outshine*
		eɪ → əʊ əʊ	*stave*
		e → əʊ +d əʊ +d	*tell, sell, foretell, retell*
2C-9	27	aɪ → ɔ: (+t) ɔ: (+t)	*buy, fight*
		i:+ ch or k → ɔ: +t ɔ: +t	*teach, seek, rethink*
		ɪ + ng or nk → ɔ: +t ɔ: +t	*bring, think*
		ʌ+ ch → ɔ: +t ɔ: +t	*catch*
2D - past = perfect +n			
2D-1	28	eə → ɔ: ɔ: +n	*wear, tear, bear, swear*
2D-2	29	aɪ → □ ɪ ɪ (+n)	*light, hide, slide, bite, chide*
2D-3	30	aɪ → □ eɪ eɪ +n	*lie-lay-lain*
2D-4	31	i: → □ əʊ əʊ +n	*speaks, eat, freeze, weave, cleave, bespeak, heave, unfreeze*
		eɪ → □ əʊ əʊ +n	*break, wake, awake*
		u: → □ əʊ əʊ +n	*choose*
2D-5	32	e → ɒ ɒ (+n)	*get, forget, beget*
		i: → ɒ ɒ +n	*tread*
Three- vowel paradigm			
3A - present ≠ past ≠ perfect			
3A-1	33	aɪ → əʊ ɪ +n	*write, drive, rise, ride, arise, rewrite, stride, strive, override, underwrite, smite, cowrite, thrive, overwrite, handwrite, bestride*
3A-2	34	i → □ æ ʌ	*begin, sing, drink, ring, sink, swim, spring, shrink, stink*

3A-3	35	aɪ → uː əʊː +n	*fly*

To find examples of all possible IFs in our sample, first we extract all possible vowels found in existing English past forms of IVs and apply them to IVs found in our sample with the use of the wildcard tool on WebCorp. The wildcard tool can help us to search for alternative characters within a word. For example, to search for possible IFs within the verb *run-ran* in the past form, we get *rin, ren, ron, roun, run* in reference to *hit, set, lost* and *found*, respectively. To reach these possible IFs, the wildcard tool of the square brackets *r[i/e/o/ou/u]n* is used. Finally, all the hits of IFs are checked manually in the contexts to remove the unwanted cases.

3.5.2 The diachronic snapshot

The way of selecting the verb sample and the procedures that will be used in the diachronic analyses of this study are mentioned in the subsection below to investigate word frequency effects on regularization and irregularization processes over time in the selected sample.

3.5.2.1 Case 5: (ir)regularization and word frequency

To investigate whether or not IVs are regularized constantly over time, I choose the sample of this diachronic analysis from the WebCorp corpus that covers the period January 1995-December 2010. This period (16 years) is divided into two-time spans: the old span (1995-2002) and the new one (2003-2010). I select the same sample of the IVs used in the question 1 (see appendix 2) for the investigation of regularization and the same IVs of the sample used in the question 3 for the investigation of irregularization (see appendix 4). Then, in the old and new spans, word frequencies of IVs split by form and frequency are collected from the selected sample. Similarly, word frequencies of RFs and IFs split by form and frequency are collected in the old and new spans from the selected sample to draw a comparison between the two spans (see appendices 8, 9, 10 and 11). By doing so, I can investigate whether verbal changes have a diachronic tendency towards regularization only (favoring the dual mechanism approach) whether irregularization processes can also be attested (favoring single mechanism approaches) (see chapter 2 and section 3.2 for more details).

In the next chapter, the selected data of this study will be analyzed in an attempt to answer the research questions mentioned in 3.2.

4 Chapter Four: Data Analysis

In this chapter, I will explore verb (ir)regularization processes synchronically and diachronically in the selected corpus as illustrated in chapter 3. First, I provide the necessary overview of the single-dual debate for analyzing the case studies of this thesis (see chapter 3). For each case study, I will display descriptive statistics that shows a general overview of the selected data. To do so, tables and different types of graphs will be used for comparative and descriptive purposes. Afterwards, I will conduct statistical models to exam the significance of the differences in the results whenever necessary before drawing conclusions.

4.1 An overview of the single-dual debate

As indicated in chapters 2 and 3, a longstanding debate in linguistics and psycholinguistics relates to how linguistic information is processed by the human mind (Chomsky and Halle 1968; Bybee 1995; Pinker 1999). In this debate, one question arises as to how morphological knowledge is mentally represented and this question is particularly relevant to work in this thesis. Two different types of approaches can be distinguished in this respect: single and dual mechanism approaches of morphological processing.

Single mechanism approaches posit no fundamental distinction between regular and irregular inflections and contend that both are built via a single mechanism. Hence, these approaches hypothesize that all morphological processes are taken care of by one single mental mechanism – either a rule system or an associative system. Proponents of rule-based models assume that both RVs and IVs are generated by rules (Chomsky and Halle 1968; Halle and Mohanan 1985). For example, RVs are generated by a rule that adds a suffix *–ed* to the stem of the verb (e.g., *walk-walked*). To generate the irregular form *ring-rang*, for example, a rule that changes a vowel from /i / to /a/ is applied when it occurs before a consonant cluster *-ng*.

By contrast, proponents of associative models assert that all inflected words are stored within a single associative system (Rumelhart and McClelland 1986; Smolensky 1996; Bybee 1995; Seidenberg and Gonnerman 2000; McClelland and Patterson 2002). According to these connectionist models of morphological processing, people store associations between the sounds of stems (input codes) and the sounds of past-tense forms (output codes) and generalize the associations to new words if they are similar to old words. So, for example, the learning of the past form of the verb *walk* can be made by storing associations between the sounds of stem *walk* as the input and the sounds of past-tense form *walked* as the output. This regular pair (*walk-walked*) reinforces the connections between the *alk* input nodes and the *alked* output nodes, and thus generalizes them to similar pairs like *talk-talked*. Hence, by means of this single mechanism of association between input and output representations, people can learn both RVs and IVs without making use of rules. Both single mechanism approaches (rules or storage) do not predict a clear correlation between (ir)regularity and word frequency, since there is no firm distinction between RVs and IVs.

Within the dual mechanism approach, the core features of the two previous models are combined. Followers of this approach claim that IVs are stored in the associative memory, while RVs are generated by rules ($V_{past} = V_{stem}$ + *-ed* as in *play-played*), applying by default upon the failure to retrieve a stored irregular from the associative memory (Pinker and Prince 1988; Marcus et al. 1995; Pinker 1999; Pinker and Ullman 2002). The dual mechanism predicts that there is a positive correlation between word frequency and irregularity: the more frequent words are, the more likely they are irregular. Pinker (1999) supports this argument stating that the top ten frequent verbs in English, *be, have, do, go, say, can, will, see, take* and *get*, are all irregular. Similarly, KuCera and Francis (1967) claim that of the top 30 most frequent English verbs in the past form 22 are irregular. Nevertheless, RVs are rule-produced and are consequently frequency-insensitive.

Diachronically, it has been attested that the number of IVs in English has steadily declined over time. IVs with low frequency are regularized more often than IVs with high frequency (Fries 1940; Lieberman et al. 2007; Michel et al. 2011). Yet, this one-sided view that looks at morphological changes in the direction of regularization has been objected by some linguists, as changes in the other direction, the direction of irregularization, have been observed as well (Nübling 2000 Peters 2009 and Fertig 2013).

This study is a contribution to this debate by exploring the possibility of a relationship between word frequency and (ir)regularity in the English verbal system by running a corpus study. I aim to investigate whether there is a relationship between:

- (ir)regularity and word frequency in a synchronic snapshot
- (ir)regularization and word frequency in a synchronic snapshot
- (ir)regularization and word frequency in a diachronic snapshot

If there is a link between irregularity and word frequency, this (in principle) would provide evidence for the dual mechanist view arguing against single mechanist views. Section 4.2 will cover the exploration of the relationship between the (ir)regularity and word frequency in the synchronic snapshot. Sections 4.3 and 4.4 will investigate the relationship between the (ir)regularization and word frequency in the synchronic snapshot. Finally, sections 4.5 and 4.6 will capture the relationship between (ir)regularization and word frequency in the diachronic snapshot.

4.2 (Ir)regularity and word frequency

The dual mechanism approach predicts a relationship between irregularity and word frequency, described in terms as a reflection of memory cost that IVs produce to be retrieved successfully from memory. Yet, RVs are generated by a rule and do not need to be fully retrieved from the associative memory. In contrast, single mechanism approaches predict that there is no clear relationship between (ir)regularity and word frequency. They affirm that there is no principled difference between RVs and IVs.

With the aim of testing these predictions, in this section, I investigate whether there is a link between (ir)regularity and word frequency. This leads us to the following question:

- Are IVs generally more frequent than RVs in Contemporary English?

While RVs form their past and perfect by adding *-ed* to the stem (e.g., *play-played-had played*), IVs form them in a number of different ways. Some verbs distinguish all forms, e.g., *write- wrote-written* or *sing-sang-sung*, but others show a two-way distinction e.g., *mean-meant-meant* or a one-way distinction, e.g., *cut-cut-cut*. To see to what extent there is a relationship between (ir)regularity and word frequency in the past and perfect, the following question is addressed:

- Are IVs generally more frequent than RVs in the past and perfect forms in Contemporary English?

To start, our preliminary analysis aims to investigate the claims of Pinker (1999) and KuCera and Francis (1967) that state that the most frequently occurring English verbs are IVs. For the purpose of this investigation, I selected the top 30 verbs with the highest word frequency in the past form from the selected sample. Table 11 below illustrates frequency distributions of these verbs in the past form from our sample. Of these 30 verbs, 21 are IVs and the remaining 9 are RVs. The first top ten verbs are all IVs that comprise 71% of the total word frequencies. Hence, the observation made by Pinker, KuCera and Francis is supported by the top 30 verbs in the past form in our sample. From the dual mechanism perspective, IVs should be sensitive of frequency in order to be stored and retrieved successfully from the associative memory.

Table 11: Frequency distributions of the top 30 verbs in the past form from the selected sample

N.	Past Verb	Word frequency	%
1	be	1,453,570	22.47%
2	say	777,450	12.02%
3	do	527,547	8.15%
4	have	454,721	7.03%
5	get	364,219	5.63%
6	make	248,679	3.84%
7	go	206,213	3.19%
8	think	196,651	3.04%
9	come	185,412	2.87%
10	take	170,648	2.64%
11	want	147,106	2.27%
12	tell	128587	1.99%
13	start	125,084	1.93%
14	see	122744	1.90%
15	find	122430	1.89%
16	write	121799	1.88%
17	use	100,311	1.55%
18	give	98449	1.52%
19	know	97106	1.50%
20	leave	87478	1.35%
21	call	82341	1.27%
22	put	79044	1.22%
23	play	76218	1.18%
24	look	75739	1.17%
25	feel	75681	1.17%
26	ask	71364	1.10%
27	win	69927	1.08%
28	love	68925	1.07%
29	post	67393	1.04%
30	lose	66879	1.03%
	Total	6,469,715	**100.00%**

As a next step, I want to make a comparison between the 250 verbs with highest word frequencies and the 250 verbs with lowest word frequencies in our sample (see chapter 3 for more details) in order to explore whether IVs are generally more frequent than RVs. To do so, I present an overall picture of word frequency distributions of the selected verbs divided by type, form and frequency, as reported in table 12 below.

Table 12: Word frequencies of the verbs by type, form and frequency in the selected sample

Type / Form	Word frequencies of the verbs		Total
	With high frequency	With low frequency	
IVs +RVs	19,762,462		
IVs	14,728,139	55,439	14,783,578
RVs	4,881,793	97,091	4,978,884
IVs / past	10,778,120	14,758	10,792,878
RVs / past	3,019,513	447	3,019,960
IVs / perfect	3,950,019	40,681	3,990,700
RVs / perfect	1,862,280	96,644	1,958,924

Table 12 displays that of the total word frequency of the selected verbs (19,762,462), word frequency of IVs (14,783,578) is higher than the one of RVs (4,978,884). As for the frequent group, word frequencies of IVs are higher than the ones of RVs in both forms (IVs: 10,778,120 for the past, 3,950,019 for the perfect versus RVs: 3,019,513 for the past, 1,862,280 for the perfect). In the infrequent group, only in the past form, word frequency of IVs (40,681) is higher than that of RVs (447). The differences in word frequencies between IVs and RVs suggest a relationship between word frequency and irregularity. In figure 13, I illustrate the data (split by type) in histograms. The skewness of frequency distributions is reduced by applying the logarithmic transformation. In the histogram of RVs, we can see that the distribution of verbs is bimodal showing modes in the bin between (0.5 and 1) and the bin between (4 and 4.5). This camel picture of RVs displays a logical distribution, as there is a large number of verbs in both frequency classes (low and high) around the two peaks. IVs, however, are more normally distributed. Here we can see that the number of IVs increases in the middle showing a single curse (around 1:5 -4:5). IVs with high and low frequency appear to be from one class. This shows us that there is a frequency effect on the distribution of the data divided by type: word frequency distributions of RVs and IVs are different.

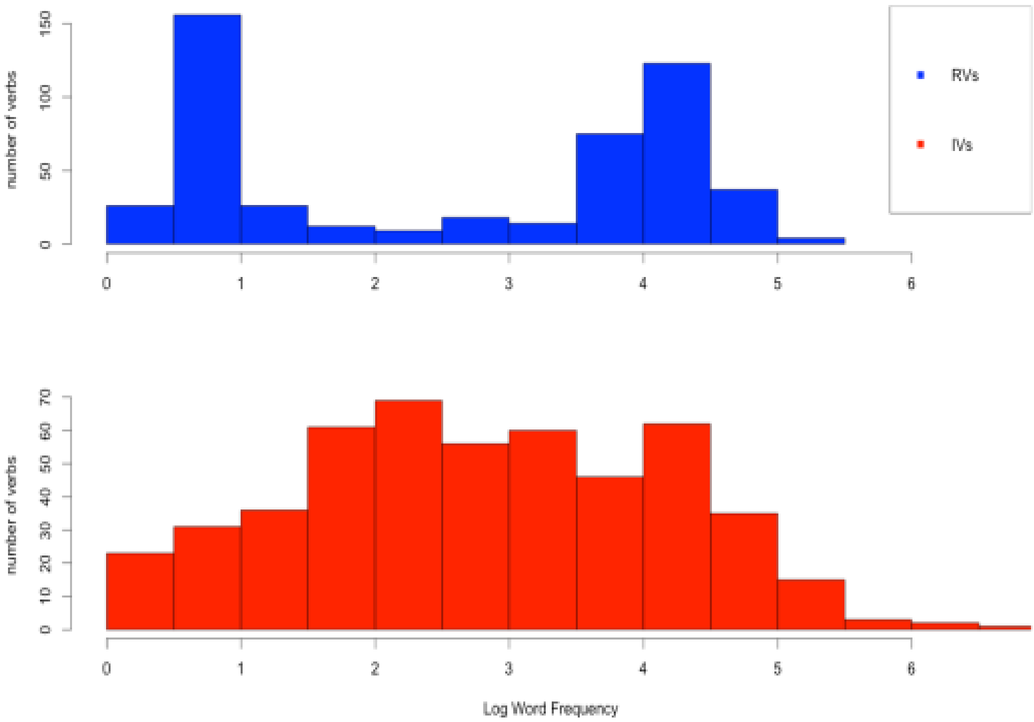


Figure 13: Histograms of word frequencies of the verbs split by type

In order to compare the distribution of the data divided by type, form and frequency, I will present side-by-side boxplots in figure 14. The boxplots of RVs are blue and the ones of IVs are red.

In the frequent group, the distributions of IVs and RVs in both forms have roughly the same center (medians of IVs: 4.0 for the past and 3.9 for the perfect versus medians of RVs: 4.2 for the past and 4.1 for the perfect). However, word frequencies of IVs have larger variability than the ones of RVs, both in terms of IQR (IVs: 1.3 for the past and 1.4 for the perfect versus RVs: 0.4 for the past and 0.7 for the perfect) and in terms of range (IVs: 3.4 for the past and 3.8 for the perfect versus RVs: 1.1 for the past and 2.1 for the perfect). Moreover, the boxplots show that the distributions of RVs and IVs are right skewed in the past form and left skewed in the perfect form showing some high values in each direction.

Comparing the boxplots of the infrequent group, we can see that medians of IVs in both forms are higher than the ones of RVs (medians of IVs: 2.0 for the past and 2.1 for the perfect versus medians of RVs: 0.5 for the past and 1.0 for the perfect). Yet, only in the past form, word frequency of IVs (IQR= 1.2 and range= 2.2) displays more variability than the one of

RVs (IQR= 0.1 and range= 0.3). Hence, the results obtained from this figure stress that, in both forms, word frequency distributions of IVs are different from the ones of RVs.

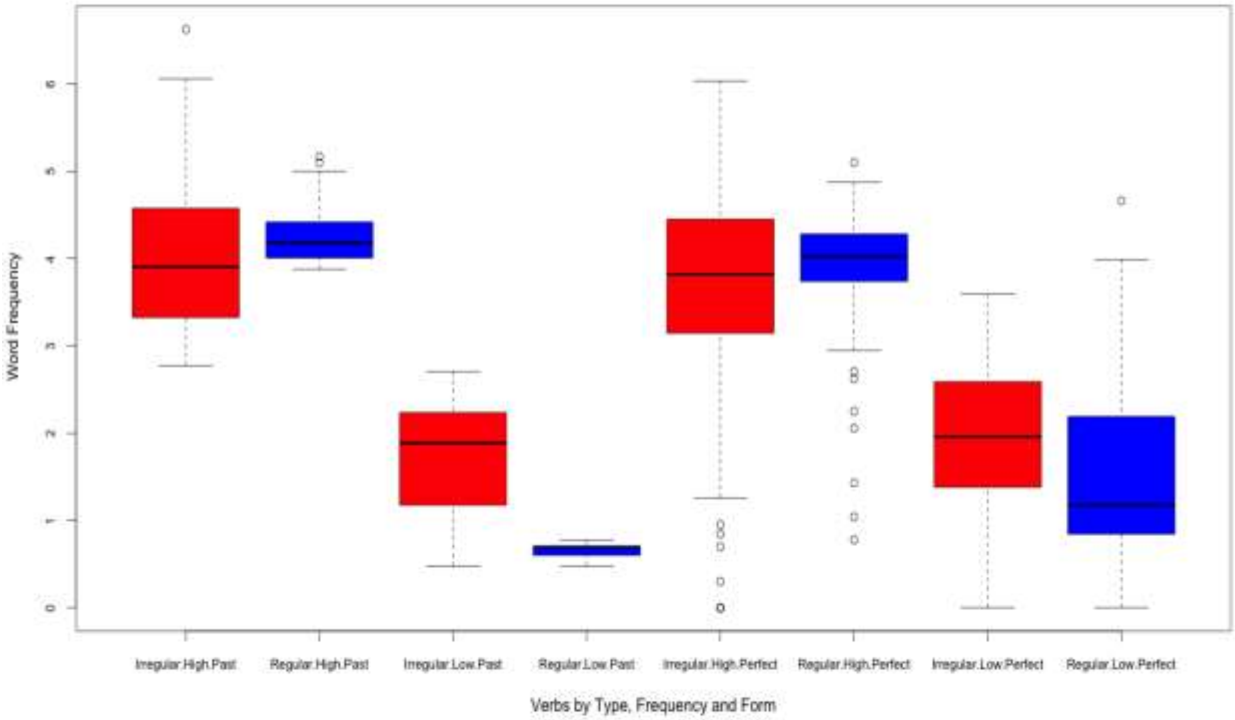


Figure 14: Box plots of word frequencies of the verbs split by type, form and frequency

Now, I aim to have a view about the central tendency of the data distribution; this informs us about the balance points in the subset distributions of the data. For this purpose, table 13 displays mean frequency of the data divided by type, form and frequency.

Table 13: Mean frequencies of the verbs by type, form and frequency in the selected sample

Type / Form	Mean frequencies of verbs		Total mean frequency
	With high frequency	With low frequency	
IVs	58,913	222	29,567
RVs	19,527	388	9,958
IVs / past	86,225	118	43,172
RVs / past	24,156	4	12,080

IVs / perfect	31,600	325	15,963
RVs / perfect	14,898	773	7,836

Table 13 gives an overview that total mean frequency of IVs is higher than that of RVs (29,567 versus 9,958). Considering the frequent group in both forms, mean frequencies of IVs are higher than the ones of RVs (IVs: 86,225 for the past, 31,600 for the perfect versus RVs: 24,156 for the past, 14,898 for the perfect). Nevertheless, in the infrequent group, only in the past form, mean frequency of IVs is higher the one of RVs (IVs: 118 versus RVs: 4).

To examine the effect of type, form and frequency on word frequency of the verbs in the selected sample, I conducted a statistical model. A linear model was adopted, where *word frequency* was considered as a dependent variable and the factors: *type* (with two levels: regular and irregular), *form* (with two levels: past and perfect) and *frequency* (with two levels: high and low) were included as fixed factors. A logarithmic transformation was applied to the data to remove most of the skewness of frequency distribution. The linear model reveals that the main effects of *type* ($\beta = 0.55$, $t = 2.39$, $p = 0.02$), *form* ($\beta = -0.85$, $t = -3.73$, $p = 0.0002$) and *frequency* ($\beta = -5.24$, $t = -22.97$, $p < 2e-16$) are significant. Moreover, the effects of the interaction between *type* and *frequency* ($\beta = -3.04$, $t = -9.44$, $p < 2e-16$), *form* and *frequency* ($\beta = 1.05$, $t = 3.25$, $p = 0.001$), and *type*, *form* and *frequency* ($\beta = 1.57$, $t = 3.45$, $p = 0.0006$) are all significant too. Instead, the effects of the interaction between *type* and *form* ($\beta = 0.08$, $t = 0.26$, $p = 0.798$) appear to be not significant. These findings confirm that the differences between frequency distributions of the verbs split by *type*, *form* and *frequency* are statistically significant. However, the differences between frequency distributions of IVs and RVs in the past and the perfect are not significant. This supports the idea that IVs occur more often than RVs. Therefore, I tentatively conclude that there is a relationship between word frequency and irregularity in Contemporary English. However, I also conclude that there is no clear relationship between word frequency and irregularity in the past and perfect forms.

To ensure that these results are uniform across different verbs and they are not due to few verbs with the highest frequency, I eliminate 4 IVs and 4 RVs with the highest word frequencies in the past form and their related perfect form from the selected sample.Table 14 displays these verbs with their word frequencies.

Table 14: Word frequencies of the verbs with the highest word frequencies in the selected sample

Type	Verbs with highest word frequency	Past	Perfect
RVs	want	147,106	8,097
	start	125,084	32,829
	use	100,311	126,113
	call	82,341	74,400
IVs	be	4,215,057	1,081,787
	have	1,147,344	103,941
	do	909,206	243,227
	say	777,450	51,786

After elimination of the verbs with the highest frequencies, total mean frequency of IVs (12,906) remains larger than the one of RVs (8,843) (see table 15). Furthermore, when considering the frequent group in both forms, mean frequencies of IVs are larger than the ones of RVs (IVs: 30,819 for the past, 20,407 for the perfect versus RVs: 21,196 for the past, 13,395 for the perfect). Focusing on the infrequent group, only in the past form, mean frequency of IVs is higher than the one of RVs (IVs: 15,462 versus RVs: 7,087).Table 16 displays that all the differences of the data distribution are statistically significant.

Hence, the relationship between word frequency and irregularity that previously observed is confirmed even after the elimination of the verbs with highest word frequencies. This supports the hypothesis that IVs are generally more frequent than RVs. However, there is no clear relationship between (ir)regularity and word frequency when comparing the past and perfect forms. This first conclusion is consistent with the predictions of the dual mechanism model claiming that irregulars are dependent on word frequency to be stored and retrieved successfully from the associative memory. Regulars, under this approach, are not sensitive to such frequency, as they are rule-produced.

Table 15: Mean frequencies of the verbs split by type, form and frequency after the elimination of the highest frequency values in the selected sample

Type / Form	Mean frequencies of the verbs		Total mean frequency
	With high frequency	With low frequency	

IVs	25,613	198	12,906
RVs	17,296	391	8,843
IVs / past	30,819	106	15,462
RVs / past	21,196	4	10,600
IVs / perfect	20,407	290	10,349
RVs / perfect	13,395	779	7,087

Table 16: The statistical results of the linear model to examine the effect of type, form and frequency on word frequency of the verbs after the elimination of the highest frequency values in the selected sample

Independent variables	β	T-value	P-value
Frequency	-5.15	-22.87	< 2e-16
Type	0.65	2.87	0.004
Form	-0.815	-3.62	0.0003
Type*Form	0.05	0.17	0.8629
Frequency*Type	-3.08	-9.68	< 2e-16
Frequency*Form	0.99	3.11	0.0019
Frequency*Type*Form	1.586	3.52	0.0004

In conclusion, the results of the synchronic analysis in this corpus study confirm the hypothesis that verbs with high frequency are more likely to be irregular. IVs are high-frequent, whereas RVs are not in our sample. Additional support to this conclusion comes from the fact that the ten most common verbs (*be, have, say, do, get, make, go, think, come* and *take*) are irregular in the selected sample. Tentatively, these results are accounted for under the dual mechanism approach and thus run against single mechanism approaches. The frequency of IVs in English has been taken to be as important evidence in favour of the dual mechanism approach, where only IVs are stored whole in the associative memory structure while RVs are generated by rules. By contrast, single mechanism approaches do not commit to this fundamental separation between RVs and IVs, and assert that both RVs and IVs are generated via a single mechanism (either rules or storage). Therefore, these approaches do not predict a relationship between (ir)regularity and high frequency.

In the next two sections (4.3 and 4.4), the relationship between word frequency and (ir)regularization in the synchronic snapshot will be explored in order to investigate whether these processes also take place in Contemporary English. If that appears to be the case, further investigation needs to be carried out to explore the diachronic relationship between word frequency and (ir)regularization. This, in turn, may generate more evidence in favour with or against single and dual mechanism approaches.

4.3 The synchronic analysis of regularization

A link between word frequency and regularization processes has been observed and examined by many researchers: IVs with low frequency are regularized more often than IVs with high frequency (Pinker 1999; Lieberman et al. 2007; Michel et al. 2011 among other). Under the dual mechanism approach, IVs with low frequency are predicted to be more disposed to regularization processes due to the weaker memory traces they possess and hence they are less resistant to regularization processes. IVs with high frequency, however, are easier to be accessed and accordingly more resistant to these processes. So, a regularization rate increases as word frequency of IVs decreases. In this section, I examine whether English undergoes regularization processes in the synchronic snapshot by investigating the following questions:

- Do regularization processes take place in Contemporary English? If so, are IVs with low frequency regularized more often than IVs with high frequency in the past and perfect forms?

Furthermore, I explore regularization processes in a limited set of doublet verbs that can be conjugated as both IVs and RFs in the English language like *burn* and *light* are conjugated as *burned/burnt* and *lighted/lit* in (For more details see section 3.3). I aim to explore whether there is a link between vowel change and regularization processes by checking to what extend presence or absence of vowel change may possibly account for the variation in such verbs**.** I Intuitively expect that verbs with no vowel change will be less resistant (and consequently display more RFs) than verbs with vowel change in these processes. De Clerck and Vanopstal (2015) explore the impact of vowel change on the retention of regularization processes in 11 doublet verbs (see more details in section 2.3.3). Their study discloses that there is a link between the salience of vowel change and regularization processes. In our study, I investigate whether this tendency can still be demonstrated in the selected sample and accordingly the following question is addressed:

- Do regularization processes occur more frequently in the cases where IVs and RFs show no vowel change?

4.3.1 Regularization and word frequency

In this section, I explore whether in Contemporary English regularization processes take place in the synchronic snapshot and whether there is a tendency for verbs to be regularized, with resistance to regularization coming as a result of high word frequency. For this purpose, I select the same verbs of the sample used in the question 1 mentioned in 3.2: 500 IVs with their word frequencies in the past and perfect forms (see appendix 2). Then, word frequencies of RFs are collected (see appendix 4 and see chapter 3 for more details). Table 17 below illustrates a general overview of word frequency distributions of IVs and RFs split by form and frequency. Moreover, relative frequencies of RFs are computed, as the number of word frequencies of RFs depends on the size of the selected sample.

Table 17: Frequency distributions of IVs and RFs in the selected sample

Type/Form	High frequency verbs		Low frequency verbs		Total	
	Word frequency	% of RFs	Word frequency	% of RFs	Word frequency	% of RFs
IVs	14,728,139		55,439		14,783,578	
RFs	64,260	0.43 %	81,670	60 %	145,930	0.98 %
IVs / past	10,778,120		14,758		10,792,878	
RFs/ past	36,348	0.34 %	46,841	76 %	83,189	0.76 %
IVs / perfect	3,950,019		40,681		3,990,700	
RFs/ perfect	27,912	0.70 %	34,829	46 %	62,741	1.55 %

The overall picture shows that regularization processes take place in Current English. Of the total word frequency of IVs in our sample (14,783,578), the instances of RFs are 145,930. This amount consists of 0.98 % of word frequencies of IVs in the selected sample.

Focusing on frequency, the total word frequency of RFs with low frequency (81,670) is larger than the one with high frequency (64,260). The regularization rate in the low frequency group (60%) is high compared to that in the high one (only 0.43 %). Similarly, in both forms, word frequencies of RFs with low frequency are larger than RFs with high frequency (low: 76 % for the past and 46 % for the perfect versus high: 0.34 % for the past and 0.70 % for the perfect). The differences in frequency distributions of RFs with high and low frequency may suggest a relationship between regularization processes and word frequency: high-frequency IVs may be more resistant to regularization processes than low-frequency IVs in the selected sample.

In figure 15, I illustrate the distributions of relative frequencies of RFs split by form and frequency in boxplots. The boxplots of RFs with high frequency are red and the boxplots of RFs with low frequency are blue. I apply a logarithmic transformation to reduce the skewness of the frequency distribution. As we can see in both forms, relative frequencies of RFs with low frequency have larger variability than those with high frequency both in terms of the interquartile range (IQR) (low: around 0.6 for both forms versus high: around 0.2 for both forms) and in terms of range (low: around 1.0 for both forms versus high: around 0.4 for both forms).

Moreover, the boxplots show that frequency distributions of RFs with high and low frequency in both forms are right skewed. The distributions of RFs with high frequency display some high values that lay above 0.5 of relative frequencies of RFs. From this analysis, it appears that the frequency effect on the data distribution of RFs with high and low frequency in both forms may be different. So again, there is a suggestion that less frequent IVs are regularized more quickly than more frequent IVs in our sample.

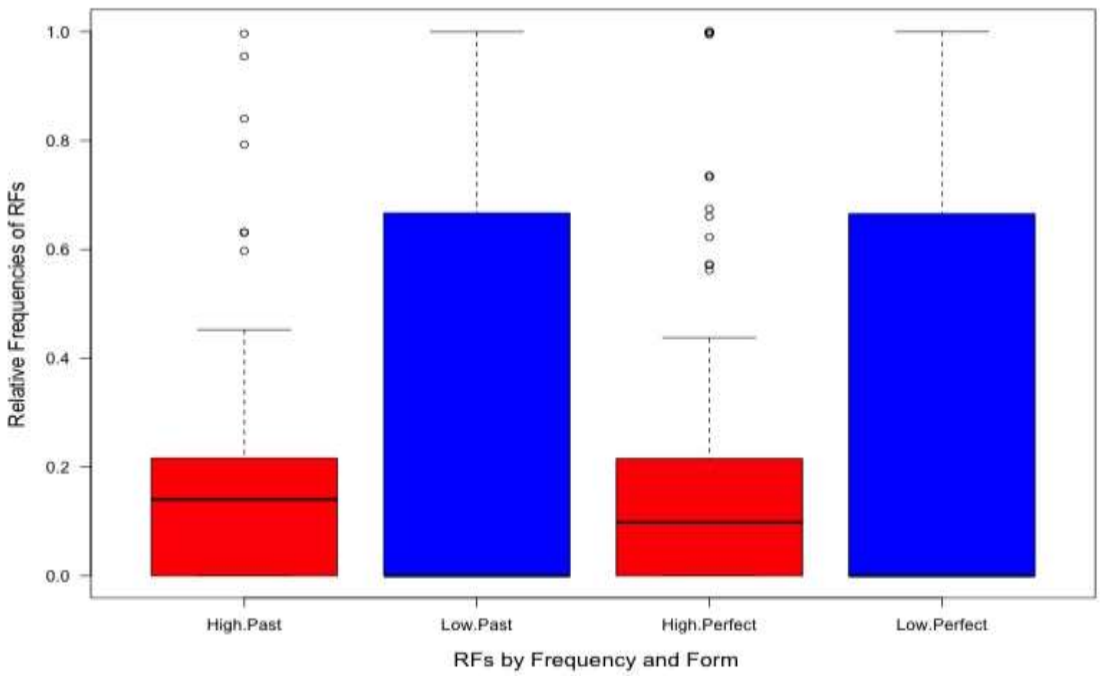
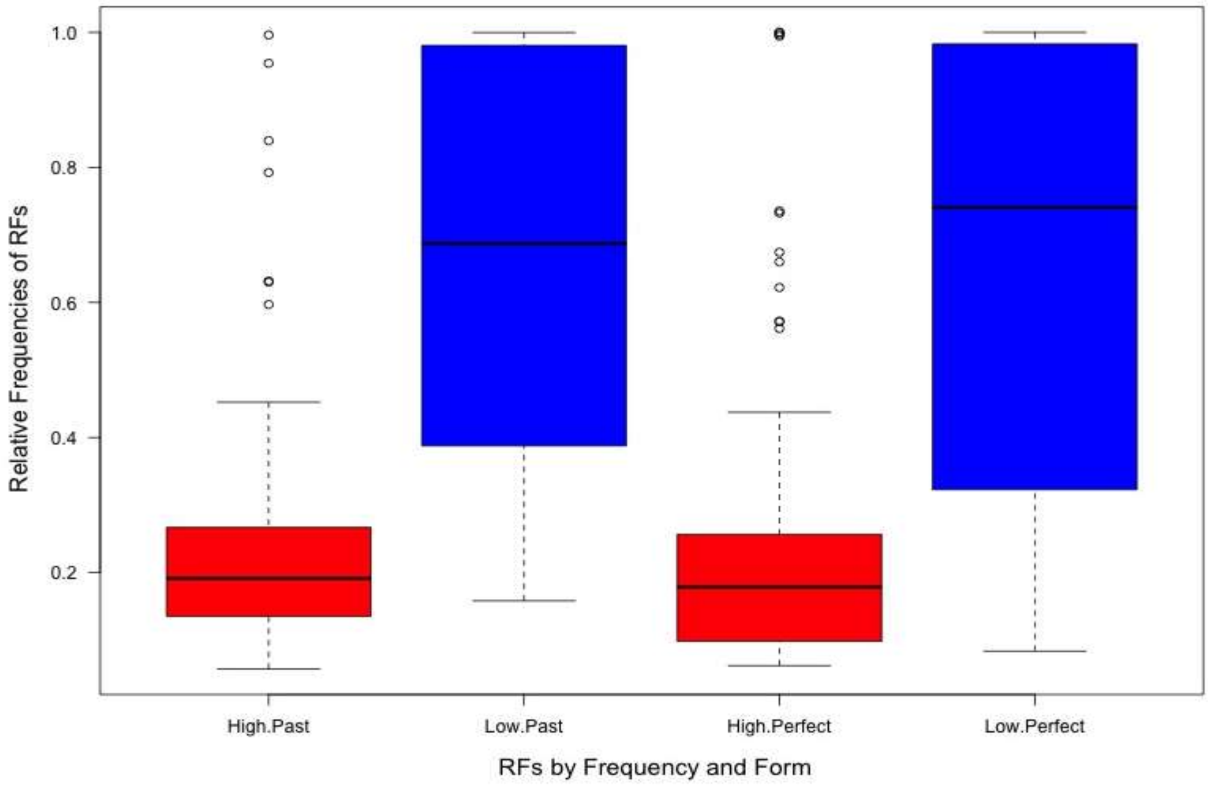


Figure 15: Box plots of relative frequencies of RFs split by form and frequency

Nevertheless, in the analysis of relative frequency, there are a lot of the verbs that are never regularized. Accordingly, many data points with zero value are displayed. The high number of zero points presented in frequency distribution can remarkably affect the shape of this data distribution. In addition, this non-normally distributed data violates one of the assumptions of a linear model[12]. In figure 16, after the removal of zero points, relative frequencies of RFs with low frequency (IQR= 0.6 and range= 1.0) still display more variability than those with high frequency (IQR= 0.3 and range= 0.5). In addition, frequency distributions of RFs with high and low frequency in both forms have different centers (low: around 0.7 versus high: around 0.2). Hence again, this tells us that frequency distributions of RFs with high and low frequency in both forms are probably different: IVs with low frequency may be regularized more often than IVs with high frequency in the past and perfect forms.



[12] In a linear model, it is supposed that variables have normal distributions. Non-normally distributed variables may falsify relationships and significance tests (For more details check: http://pareonline.net/getvn.asp?n=2&v=8).

Figure 16: Box plots of relative frequencies of RFs (without zeros) split by form and frequency

To get statistical results that are not affected by skewness of the data distribution, I conducted the following analysis of relative frequencies of RFs after removing zero points. I conducted a statistical model to investigate the effects of form and frequency on relative frequencies of the verbs in the selected sample. A linear mixed model was adopted, where *relative frequency* was considered as a dependent variable and the factors: *form* (with two levels: past and perfect) and *frequency* (with two levels: high and low) were included as fixed factors. A logarithmic transformation was applied to the data to remove most of the skewness of frequency distribution. The linear model reveals that the main effect of *frequency* ($\beta = 0.16$, $t = 4.97$, $p = 1.06e-06$) is highly significant. Instead, the main effect of *form* ($\beta = 0.02$, $t = 1.30$, $p = 0.20$) and the effect of interaction between them ($\beta = -0.02$, $t = -0.67$, $p = 0.50$) are not significant. These findings confirm that the differences between frequency distributions of RFs with high and low frequency are statistically significant. However, the differences between frequency distributions of RFs in the past and the perfect are not significant.

Summing up, on the basis of the information collected in this synchronic study, I conclude that there is a relationship between word frequency and regularization processes in our sample. Hence, IVs with low frequency are generally regularized more often than IVs with high frequency in Contemporary English. I also conclude that there is no clear relationship between word frequency and regularization in the past and perfect forms. These findings are much in line with the predictions of the dual mechanism approach in which IVs with low frequency are predicted to be more prone to regularization processes. This follows from the hypothesis that word frequency reinforces the memory representations of IVs and accordingly makes them easier to be accessed and less to be regularized.

In the next section, I will explore the salience of vowel change that is involved in the transfer from IVs to RFs as a factor that may have an influence on the retention of regularization processes.

4.3.2 The salience of vowel change and regularization

In this section, I explore the salience of vowel change that may have an effect on the retention of regularization processes in a set of doublet verbs whose past and perfect forms allow both IVs and RV like *learned/learnt*. For the purpose of this investigation, 42 doublet verbs were selected from a list of 616 English IVs[13]. In table 18, I list the 42 doublet verbs: 21 of them show no vowel change between IVs and RFs and other 21 undergo vowel change. Then, I collect word frequencies of these verbs split by type, form and vowel change in the selected sample to draw comparisons (see table 19). I intuitively assume that regularization processes that involve vowel change from IVs to RVs will meet more resistance than the ones with no vowel change

Table 18: The selected 42 doublet verbs from a list of 616 English IVs

	Without Vowel Change	Vowel Change
1.	bend-bent-bent	abide-abode-abidden
2.	bless-blest-blest	alight-alit-alit
3.	burn-burnt-burnt	beseech-besought-besought
4.	bust-bust-bust	chide-chid-chid
5.	clap-clapt-clapt	clothe-clad-clad
6.	dwell-dwelt-dwelt	creep-crept-crept
7.	forecast-forecast-forecast	dream-dreamt-dreamt
8.	geld-gelt-gelt	grind-ground- ground
9.	gird-girt-girt	hang-hung-hung
10.	ken-kent-kent	heave-hove-hove
11.	learn-learnt-learnt	kneel-knelt- knelt
12.	pen-pent-pent	leap-leapt-leapt
13.	rid-rid-rid	light-lit-lit
14.	smell-smelt-smelt	shine-shone- shone
15.	spell-spelt-spelt	slink-slunk-slunk
16.	spill-spilt-spilt	sneak-snuck-snuck
17.	spoil-spoilt-spoilt	speed-sped-sped
18.	strip-stript-stript	stave-stove-stove
19.	sweat-sweat-sweat	strive-strove-striven
20.	wed-wed-wed	tread-trod-trodden
21.	wet-wet-wet	weave-wove-woven

Table 19: Word frequencies of the 42 doublet verbs split by type, form and vowel change in the selected sample

The doublet verbs with no vowel change	The doublet verbs with vowel change

[13] This is a comprehensive list of 616 English IVs, including their base form, past simple, perfect and definitions from UsingEnglish.com.

Verbs	Past		Perfect		Verbs	Past		Perfect	
	IVs	REGs	IVs	REGs		IVs	REGs	IVs	REGs
learn	1306	31643	1415	24408	hang	6213	15	3116	405
pen	336	175	0	2044	light	3665	4	3087	209
sweat	232	12	0	249	sneak	1098	485	383	1
forecast	323	55	85	15	dream	840	1923	35	1513
spell	159	92	182	3094	grind	579	73	589	12
bust	136	1308	1	1327	leap	544	500	3	67
rid	116	4	479	3	creep	455	233	472	427
dwell	89	46	47	25	shine	452	446	260	269
spoil	85	1311	375	0	speed	425	75	246	5
spill	79	1147	172	707	weave	262	198	1453	126
ken	61	0	111	2	kneel	224	41	44	18
wet	50	4	0	15	strive	118	62	33	107
burn	20	2328	1204	5399	tread	97	42	136	9
smell	12	1742	260	4	slink	57	46	25	10
geld	6	2	11	16	abide	33	22	0	19
wed	7	0	653	313	heave	10	198	9	49
gird	5	39	7	2	beseech	9	10	0	1
bless	3	56	52	17417	chide	6	174	6	85
strip	0	1202	0	2494	alight	1	23	0	0
clap	0	245	0	142	clothe	0	14	841	740
bend	0	15	0	5	stave	0	77	0	18

Table 20 below presents the total word frequencies and mean frequencies of IVs and RFs of the 42 verbs in the past and perfect forms in the selected sample to show a general overview of the distribution of these doublet verbs. Comparing IVs, the frequencies of the verbs with vowel change (word frequency: 25,826 and mean frequency: 1,230) are higher than those with no vowel change (word frequency: 8,079 and mean frequency: 385). On the contrary, when comparing RFs, the frequencies of RFs with no vowel change (word frequency: 99,107 and mean frequency: 4,719) are higher than those with vowel change (word frequency: 8,751 and mean frequency: 417).

Table 20: Frequency distributions of the 42 doublet verbs split by form, type and vowel change in the selected sample

Type / Form	Word frequencies of the verbs

			With no vowel change	With vowel change
IVs			8,079	25,826
RFs			99,107	8,751
IVs / past			3,025	15,088
RFs/ past			41,426	4,661
IVs / perfect			5,054	10,738
RFs/ perfect			57,681	4,090
Total	IVs	Mean freq.	385	1,230
		%	25	75
	RFs	Mean freq.	4,719	417
		%	92	8

In figure 17, I illustrate frequency distribution of the data by presenting bar charts of relative frequencies of IVs and RFs in vowel change group (in red) and in no vowel change group (in blue), since the number of word frequency depends on the size of the selected sample. The regularization rates with no vowel change in both forms (past: 93% and perfect: 92%) are larger than the ones with vowel change (past: 24% and perfect: 28%).

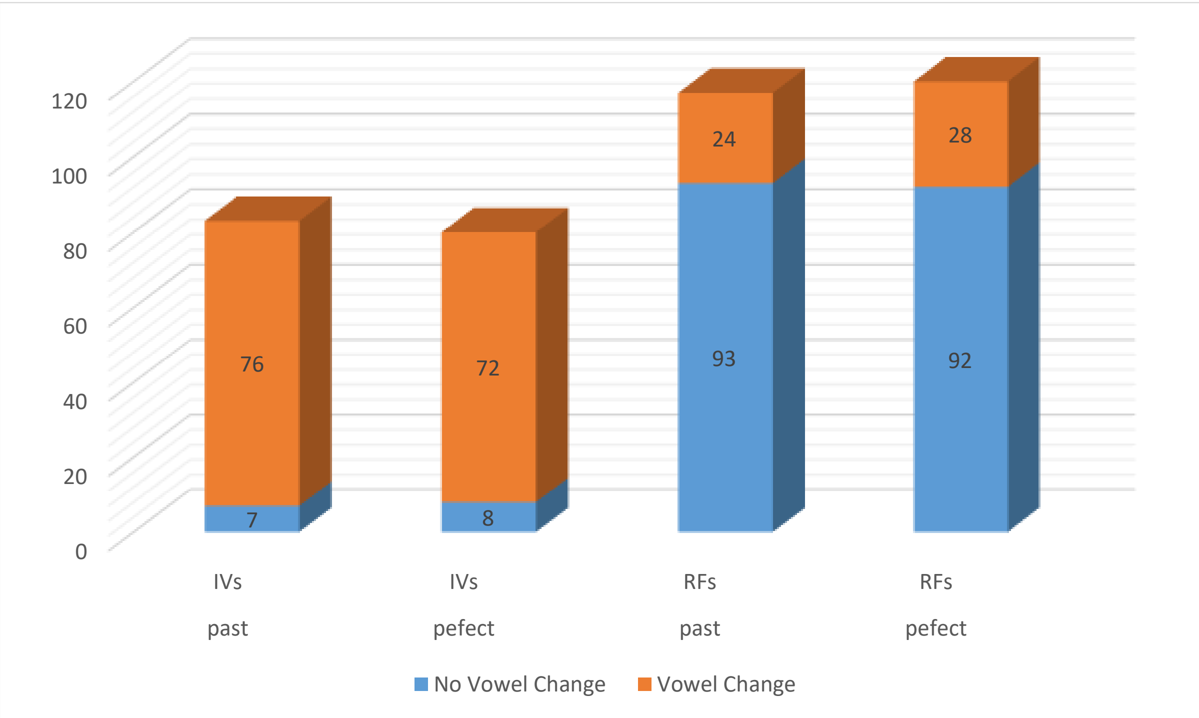


Figure 17: Bar charts of relative frequencies of IVs and RFs of the 42 verbs in the past and perfect forms

To examine the effect of the salience of vowel change and form on word frequencies of RFs in the selected sample, I conducted a statistical model. I adopted a linear mixed model, where *word frequency* of RFs was considered as a dependent variable and the variables: *the salience of vowel change* (with two levels: vowel change and no vowel change) and *form* (with two levels: past and perfect) were integrated as fixed factors. Due to few data points in the selected sample, the linear mixed model reveals that the main effects of *the salience of vowel change* (β = -0.31, t =-0.41, p =0.68) and *form* (β =0.30, t =0.51, p =0.61), in addition to effects of the interaction between them (β =-0.69, t =-0.82, p =0.41) are all not significant. Still, the findings of the descriptive analysis show that the overall regularization rate with no vowel change (92%) is higher than the one with vowel change (8%). This suggests that the salience of vowel change may have an effect on the retention of regularization processes in doublet verbs.

To have a closer look at preferences of the individual verbs for regularization processes in the selected sample, in table 21, I draw comparisons among these verbs in the two groups (with and without vowel change). I list the verbs according to their relative frequencies of RFs (high to low) in the past and perfect forms. The highlighted verbs indicate a preference of more than 50 %; the orange highlighted colour for the verbs with no vowel change and the blue highlighted colour for the verbs with vowel change. In the past form, 11 verbs are highlighted in the no vowel change group, whereas only 8 verbs are found in the vowel change group. Similarly, in the perfect form, there are 13 highlighted verbs in the no vowel change group, while only 9 verbs are observed in the vowel change group. Hence, in both forms, the regularization rates of the verbs in the vowel change group are lower than the ones in the no vowel change group. As predicted, this tells that the verbs with vowel change will meet with more resistance in regularization processes (and consequently display fewer RFs) than the verbs with no vowel change. Therefore, the findings of this descriptive analysis explain the influence of the salience of change on the retention of regularization processes in both forms: IVs with no vowel change appear to be regularized more often than IVs with vowel change Contemporary English. This is in line with results of the study of De Clerck and Vanopstal (2015) in which a correlation between regularization processes and the inflectional variation is attested.

Table 21: Preferences of the 42 doublet verbs for regularization processes

	No vowel change verbs			Vowel change verbs	
	past	perfect		past	perfect
	strip	pen		clothe	abide
	clap	sweat		stave	beseech

	bend	wet		chide	stave
	burn	bless		alight	dream
	smell	strip		heave	leap
	learn	clap		dream	chide
	bless	bend		beseech	heave
	spoil	bust		shine	strive
	spill	learn		leap	shine
	bust	spell		slink	creep
	gird	burn		weave	clothe
	spell	spill		abide	kneel
	pen	geld		creep	slink
	dwell	dwell		strive	hang
	geld	wed		sneak	weave
	forecast	gird		tread	light
	wet	forecast		speed	tread
	sweat	ken		kneel	grind
	rid	smell		grind	speed
	ken	rid		hang	sneak
	wed	spoil		light	alight

In conclusion, the data yielded by the synchronic study of regularization in sections 4.3.2 and 4.3.3 provides evidence that regularization processes take place in Contemporary English. In addition, a relationship between word frequency and regularization is attested: the regularization rate among IVs with low frequency (60%) is higher than the one with high frequency (0.43%). However, the data suggests that there is no clear relationship between word frequency and regularization in the past and perfect forms. These findings are consistent with the predictions of the dual-mechanism approach. This approach hypothesizes that word frequency strengthens the representations of IVs in the associative memory that make them easier to be accessed and therefore more resistant to regularization processes. Hence, from the dual-mechanism perspective, IVs with low frequency are more disposed to regularization processes than IVs with high frequency.

Moreover, I had a deeper look at the data of the selected 42 doublet verbs to check whether there is a link between the salience of vowel change and regularization processes. The results of the descriptive analysis display that vowel change (or absence of it) in these verbs may suggest an effect on the retention of regularization processes: the verbs without vowel change show less resistance to regularization processes and thus be regularized faster than those with vowel change in both forms. These results are consistent with results of the study of De Clerck and Vanopstal (2015) in which a correlation between regularization processes and the inflectional variation is attested.

In the next section, I will investigate the synchronic relationship between word frequency and irregularization processes to provide further evidence either with or against single and dual mechanism approaches.

4.4 The Synchronic analysis of irregularization

Supporters of the dual mechanism approach predict that there is a relationship between low word frequency and regularization (Pinker and Prince 1988; Marcus et al. 1995; Pinker 1999; Pinker and Ullman 2002). From this dual mechanism perspective, IVs should elicit frequency effects due to their hypothesized full-form storage in the associative memory, whereas RVs are produced by means of a rule that is supposedly frequency-insensitive. Consequently, IVs with low frequency are regularized more often because of their weak representations in the associative memory and hence language changes are predicted to be in principle unidirectional, towards regularization only. Additionally, it is predicted that some instances of irregularization may appear, but they are quite rare (only 0.2% of IVs) and display phonological neighbourhood effects (Xu and Pinker 1995).

Nevertheless, single mechanism approaches predict that language changes are bidirectional, directing towards both regularization and irregularization. They hypothesize that all morphological processes are taken care of by one single mental mechanism– either a rule system or an associative system (Chomsky and Halle 1968; Halle and Mohanan 1985; Rumelhart and McClelland 1986; Bybee 1995, yang 2002; Peters 2004; and Fertig 2013). Accordingly, they propose that similar mechanisms underlie the production of both regularization and irregularization. Several of originally RVs have become irregular in English (e.g. *sneak-snuck, drag-drug, hang-hung, string-strung, stick-stuck, strike-struck, dig-dug, wear-wore, wake-woke, light-lit, ring-rang, catch-caught, kneel-knelt, make-made* and *cost-cost*). Fertig (2013) argues that these IFs are necessary to a balanced understanding of the history of English irregulars. Furthermore, followers of connectionist models predict that instances of irregularization exhibit phonological similarity because of their hypothesis of storage in the associative memory. They attest irregularization rates to be between 3% and 24% of IVs (Rumelhart and McClelland 1986; Plunkett and Marchman 1991; Sproat 1992). These irregularization rates are rejected by followers of the dual mechanism approach claiming that they are much fewer in reality.

The findings illustrated in section 4.3.2 suggested that verbal changes are synchronically unidirectional in Contemporary English, towards regularization with a rate of 0.98%. In the next section, I investigate whether irregularization processes can synchronically be attested as well. I aim to explore whether these verbal changes are bidirectional, towards both regularization and irregularization, in Contemporary English. If that appears to be the case, this will form evidence in favour of single mechanism approaches and thus speak against the dual mechanism approach.

4.4.1 Irregularization and word frequency

Here, I explore whether there are instances of irregularization in the selected sample mentioned below and if so to what extent irregularization and word frequency are related. If irregularization processes are attested in our sample, this provides evidence that the current movements in English verbal system are bidirectional favouring single mechanism approaches. Accordingly, the following questions are addressed:

- Do irregularization processes take place in Contemporary English?
- If so, is there a relationship between irregularization and word frequency in the past and perfect forms?

For the purpose of this investigation, I select the same IVs of the sample used in the question 2: 500 IVs in the past and perfect with their word frequencies, excluding suppletives[14] (see appendix 5). Then, word frequencies of their corresponding IFs in both forms are collected from the selected sample (see appendices 6 and 7 and for more details see chapter 3).

To have a general overview of the data distribution in the selected sample, word frequencies of IVs and IFs split by form and frequency are reported in table 22. In addition, percentages of IFs have been calculated, as word frequencies of IFs depend on the size of the selected sample.

Table 22: Frequency distributions of IVs and IFs in the selected sample

Type/ Form	High frequency verbs	Low frequency verbs	Total

[14] The suppletives have to be excluded, as I have to test to what extent certain classes of IVs are applied to IVs of other classes and it is hard to classify these suppletives (with little or no correlation between their past and perfect forms, e.g., be, have, do and go) into any class.

	Word frequency	% of IFs	Word frequency	% of IFs	Word frequency	% of IFs
IVs	6,737,521		51,121		6,788,642	
IFs	4,748	0.07%	873	2%	5,621	0.08%
IVs / past	4,299,434		14,213		4,313,647	
IFs/ past	1,257	0.03%	838	6%	2,095	0.05%
IVs/ perfect	2,438,087		36,908		2,474,995	
IFs/ perfect	3,491	0.14%	35	0.09%	3,526	0.12%

Table 22 shows that there are indeed instances of irregularization in the selected sample. From the total word frequency of IVs (6,788,642), word frequency of IFs are 5,621. The irregularization rate is 0.08 % of IVs. This irregularization rate is less than the regularization rate counted in section 4.3.2 (0.98% of IVs). Focusing on frequency, word frequency of IFs with high frequency (4,748) is higher than the one with low frequency (873). Nevertheless, relative frequency of IFs with high frequency (0.07% of IVs) is lower than the one with low frequency (2% of IVs). Likewise, only in the past form, relative frequency of IFs with high frequency is lower than the one with low frequency (high: 0.03 % of IVs versus low: 6% of IVs).

I conducted a statistical model to explore the effects of frequency and form on word frequencies of the verbs in the selected sample. A linear mixed model was adopted, where *word frequency* was considered as a dependent variable and the factors: *form* (with two levels: past and perfect) and *frequency* (with two levels: high and low) were included as fixed factors. To remove most of the skewness of frequency distribution, a logarithmic transformation is applied to the data. The results of the model disclose that the main effects for *form* (β =-0.04, t = -0.41, p = 0.68) and *frequency* (β = -0.07, t = -0.66, p = 0.50), in addition to the effects of the interaction between them (β = -0.06, t = -0.41, p = 0.67) are not significant. The lack of these effects suggests that neither *form* nor *frequency* plays a significant role in irregularization processes.

Hence, the results of the synchronic analysis of this study show us that there are instances of irregularization in the selected sample, but at a low rate (0.08 % of IVs). Furthermore, they provide confirmatory evidence that there is no clear relationship between high word frequency and irregularization in both forms. Therefore, the conclusion to be drawn from this analysis is that the verbal changes in Current English are not bidirectional. These results are compatible with the dual mechanism approach and thus arguing against single mechanism approaches. The dual mechanism approach, unlike single-mechanism approaches, proposes that distinct mechanisms underlie the production of regularization and irregularization. From the dual mechanism perspective, language change in this respect should be primarily unidirectional, in the direction of regularization and this has been confirmed in the analysis of section 4.3.2. Nevertheless, instances of irregularization can rarely occur and this what the results of the analysis in this section shows.

In the next step, in order to have a deeper understanding of irregularization processes, I have a closer look at all possible changes in the selected IVs classified into the 35 classes in our sample (see table 10 and for more details see chapter 3). I aim to explore whether the changed classes of IVs exhibit any phonological neighbourhood effects because of being stored in the associative memory as predicted by dual-connectionist models. IVs are expected to exhibit phonological similarity, as they are stored in the associative memory in which the activation of a word form simultaneously activates all word forms that share one or more of phonological features of the word. For example, the activation of *ring-rang* must reinforce the memory traces of neighbouring IVs like *sing-sang* or *spring-sprang* (Pinker 1999; Ullman 2000; Rumelhart and McClelland 1986). Table 23 displays which of the 35 classes are changed and which are not. Of the 35 classes, 18 show stability (no class change), while 6 display different kinds of irregularization (For more details see appendices 6 and 7). The 6 changed classes are:

1. 1A-1 (like, *put-put-put*) is changed into 2C-2 (like, *sit-sat-sat*)
2. 2C-1 (like, *cling-clung-clung*) is changed into 3A-2 (like, *ring-rang-rung*)
3. 2C-9 (like, *buy-bought-bought*) is changed into 2C-1 (like, *cling-clung-clung*)
4. 2D-4 (like, *speak- spoke-spoken*) is changed into 2C-6 (like, *meet-met-met*)
5. 2D-5 (like, *get-got-got*) is changed into 2C-2 (like, *sit-sat-sat*)
6. 3A-2 (like*, ring-rung-rung*) is changed into 2C-1 (like, *cling-clung-clung*)

Table 23: The changed and unchanged classes of IVs in the selected sample

N.	Class	Ex.	Appearance	Class-change	Ex.
1	1A-1	put-put-put	Yes	2C-2	sit-sat-sat
2	1B-1	beat-beat-beaten	No	-	-
3	1C-1	learn-learned-learned	No	-	-
4	1C-2	spoil-spoilt-spoilt	No	-	-
5	1C-3	clap-clapt-clapt	No	-	-

6	1C-4	bless-blest-blest	No	-	-
7	1C-5	make-made-made	No	-	-
8	1C-6	spend-spent-spent	No	-	-
9	1C-7	pay-paid-paid	No	-	-
10	2A -1	come-came-come	No	-	-
11	2A -2	run-ran-run	No	-	-
12	2B -1	take-took-taken	No	-	-
13	2B -2	give-gave-given	No	-	-
14	2B -3	know-knew-known	No	-	-
15	2B -4	draw-drew-drawn	No	-	-
16	2B -5	see-saw-seen	No	-	-
17	2B -6	eat-ate-eaten	No	-	-
18	2B -7	fall-fell-fallen	No	-	-
19	2C-1	cling-clung-clung	Yes	3A-2	ring-rang-rung
20	2C-2	sit-sat-sat	No	-	-
21	2C-3	stand-stood-stood	No	-	-
22	2C-4	hear-heard-heard	No	-	-
23	2C-5	find-found-found	No	-	-
24	2C-6	meet-met-met	No	-	-
25	2C-7	lose-lost-lost	No	-	-
26	2C-8	tell-told-told	No	-	-
27	2C-9	buy-bought-bought	Yes	2C-1	cling-clung-clung
28	2D-1	wear-wore-worn	No	-	-
29	2D-2	hide-hid-hidden	Yes	2D-2	light-lit-lit
30	2D-3	lie-lay-lain	No	-	-
31	2D-4	speak-spoke-spoken	Yes	2C-6	meet-met-met
32	2D-5	get-got-got	Yes	2C-2	sit-sat-sat
33	3A-1	write-wrote-written	No	-	-
34	3A-2	ring-rang-rung	Yes	2C-1	cling-clung-clung
35	3A-3	fly-flew-flown	No	-	-

Out of the selected 488 IVs, only 20 different data points of irregularization in 6 changed classes are obtained. 14 IFs are in the past form (i.e. *shit, spit, fling, wring, swing, bring, cleave, beget, sing, ring, sink, spring, shrink, stink*), whereas only 6 are in the perfect form (i.e. *spit, strike, bring, bite, cleave, forget*). Word frequencies and percentages of the 20 IVs and IFs are mentioned in table 24 below to help us draw comparisons. In figure 18, bar charts of relative frequencies of 14 IFs in the past form are displayed. The last 3 IVs namely *sink*, *beget* and *stink* present the highest relative frequencies of IFs in this group (32.2%, 79%, 79.8% respectively). Similarly, bar charts of relative frequencies of 9 IFs in the perfect form is displayed in figure

19. *Split* and *cleave* show the highest relative frequencies of IFs compared to the other IFs in this group (30.7%, 90.9% respectively).

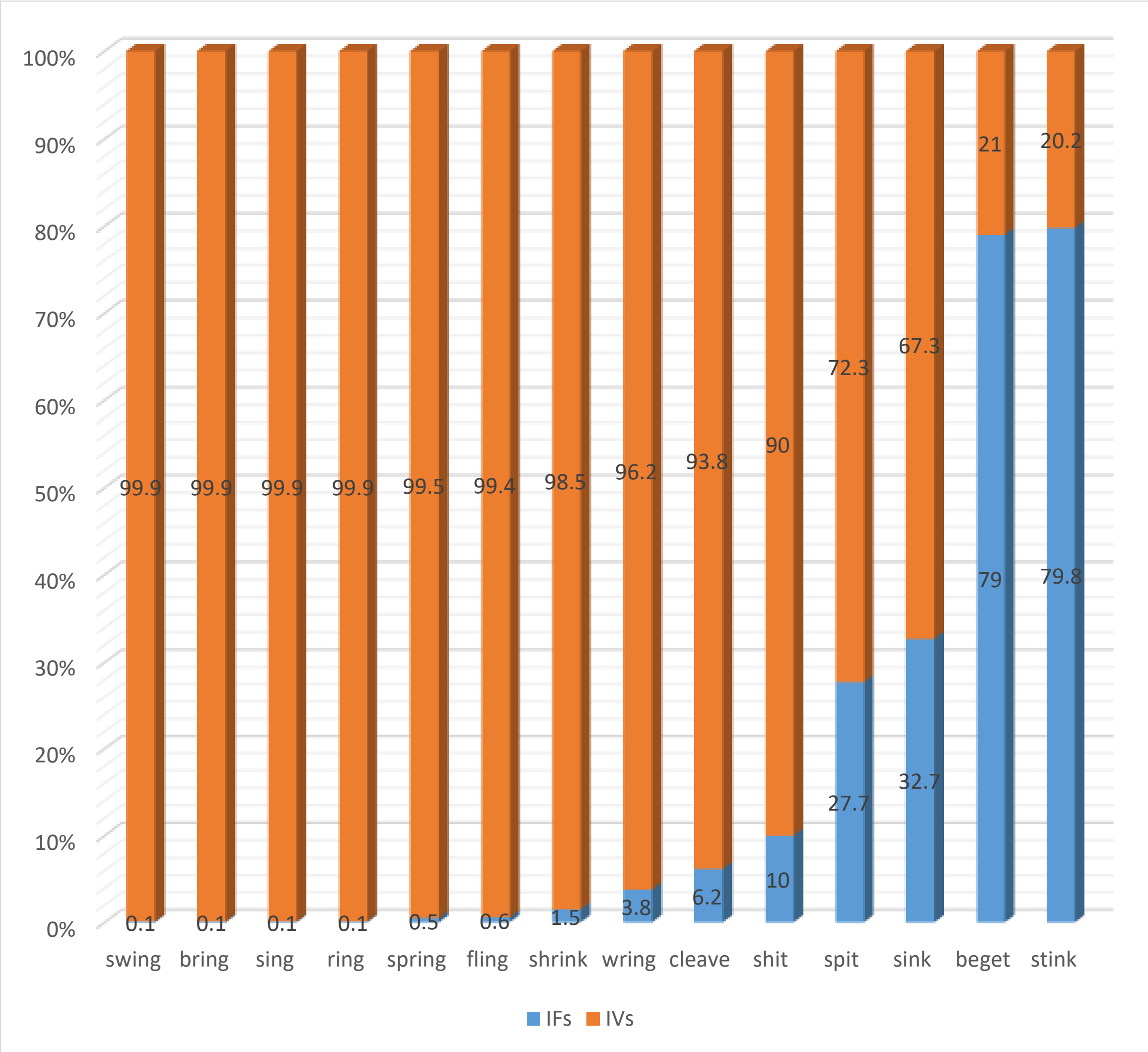


Figure 18: Bar charts of relative frequencies of 14 IFs in the past form from the selected sample

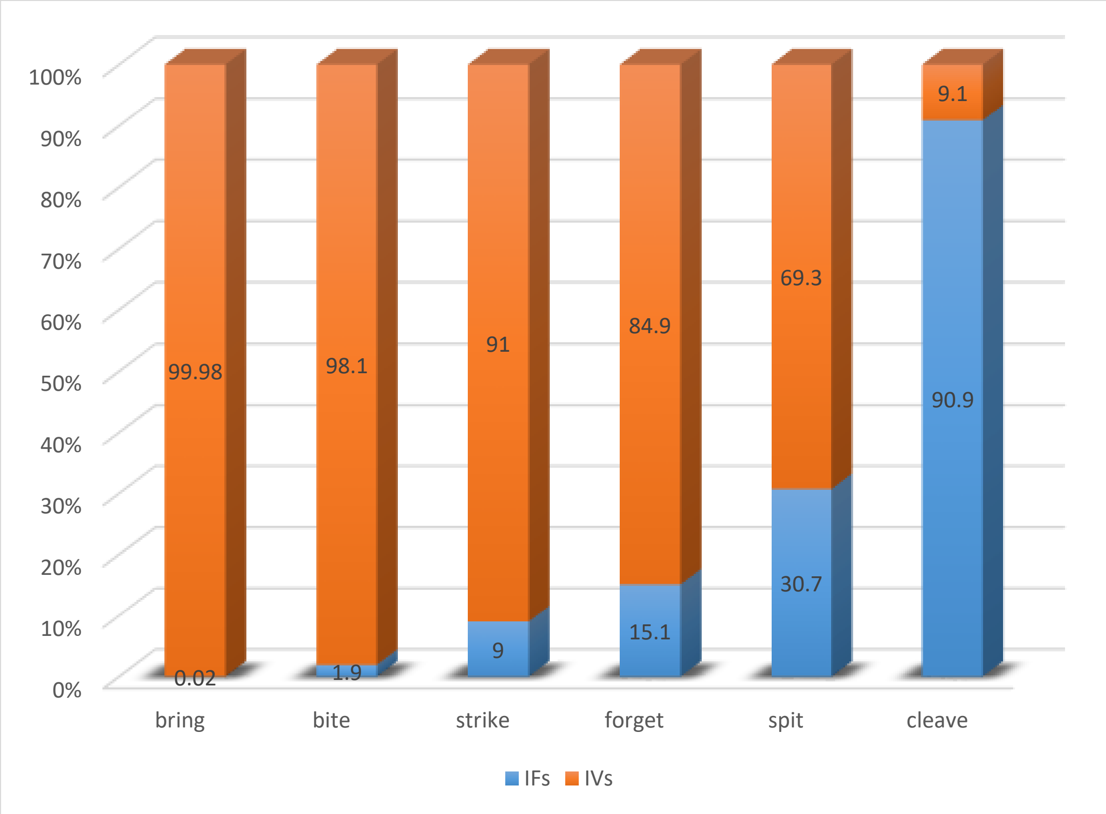


Figure 19: Bar charts of relative frequencies of 6 IFs in the perfect form from the selected sample

Table 24: Word frequencies and percentages of the 20 IVs and IFs in the selected sample

Class	IVs	Form	Word Freq.	%	Class-change	IFs	Word Freq.	%
One-vowel paradigm								
1A-1	shit	past	2,150	90	2C-2	shat	239	10
	spit	past	1,026	72.3		spat	394	27.7
	spit	perfect	321	69.3		spat	142	30.7
Two-vowel paradigm								
2C-1	flung	past	307	99.4	3A-2	flang	2	0.6
	wrung	past	50	96.2		wrang	2	3.8
	swung	past	1,675	99.9		swang	1	0.1
	struck	perfect	5,538	91	3A-1	stricken	546	9
2C-9	brought	past	40,364	99.9	2C-1	brung	27	0.1
	brought	perfect	25,991	99.98		brung	5	0.02
2D-2	bitten	perfect	1,265	98.1	2D-2	bit	25	1.9
2D-4	clove	past	76	93.8	2C-6	cleft	5	6.2
	cloven	perfect	1	9.1		cleft	10	90.9
2D-5	forgotten	perfect	15,712	84.9	2C-7	forgot	2,798	15.1

	begot	past	41	21	2C-2	begat	154	79
Three-vowel paradigm								
3A-2	sang	past	5,460	99.9	2C-1	sung	3	0.1
	rang	past	2,279	99.9		rung	2	0.1
	sank	past	1,212	67.3		sunk	588	32.7
	sprang	past	603	99.5		sprung	3	0.5
	shrank	past	271	98.5		shrunk	4	1.5
	stank	past	170	20.2		stunk	671	79.8
Total word frequency			104,512				5,621	

Table 24 displays the individual IVs that change their classes:

- from 3A-2 into 2C-1 in the past form, 6 verbs (*sing, ring, sink, spring, shrink, stink*)
- from 2C-1 into 3A-2 in the past form, 3 verbs (*fling, wring, swing*)
- from 2C-9 into 2C-1 in both forms, 1 verb (*bring*)
- from 2D-4 into 2C-6 in both forms, 1 verb (*cleave*)
- from 2D-5 into 2C-2 in the perfect form, 1 verb (*forget*). Also from 2D-5 into 2C-7 in the past form, 1 verb (*beget*)
- from 1A-1 into 2C-2 in the past form, 2 verbs (*shit, spit*). Also from 1A-1 into 2C-2 in the perfect form, 1 verb (*spit*)

In addition, within the same class (2D-2), 1 verb (*bit*) displays simplification in the perfect form (*bite-bit-bitten* into *bite-bit-bit*).

As only 20 data points of irregularization were found, I cannot evaluate the significance of the differences between these few data statistically. For this sample, I can only provide a further descriptive analysis in the next steps.

Focusing on the one-vowel paradigm, only one class change is attested; 1A-1 (*cut-cut-cut*) into 2C-2 (*sit-sat-sat*). Although I find only 3 verbs here, these verbs display considerably high irregularization rates (past: *shit* (10%), *spit* (27.7%) and perfect: *spit* (30.7%)) (see table 24).

Within the two-vowel paradigm, table 24 displays that 4 different class changes are attested (2C-1 into 3A-2, 2C-9 into 2C-1, 2D-4 into 2C-6 and 2D-5 into 2C-2). The common direction of form substitution in this paradigm is for the past form to be used for the perfect form, as in:

1. Destiny has **bit** him with sharp pointy teeth.[15]
2. I had **forgot** to put the window itself back into the frame after returning home the night before.[16]

However, in the three-vowel paradigm, the only direction of form substitution is for the perfect form to be used for the past form. Hence, the trend is towards changing class 3A-2 (i-a-u) into class 2C-1 (i-u-u) (see table 24). Here, 6 verbs undergo this type of change (*sang-sung, rang-rung, sank-sunk, sprang-sprung, shrank-shrunk, stank-stunk*), as in:

3. She **sunk** and died a Martyr to Excesive Sensibility.[17]
4. Mad Men **stunk** last night![18]

The verb *stink* shows the strongest tendency of class change from the three-vowel paradigm (*stink-stank-stunk*) into the direction of the two-vowel paradigm (*stink-stunk-stunk*). This verb gets the highest percentage (79.8% of IVs) compared to other data points of irregularization (see figure 18). This suggests a class shift from 3A-2 (i-a-u) into 2C-1 (i-u-u) for this verb. The reason behind this type of class change may lie in the fact that these 2 classes share phonological features: both have the same vowel in the present /i/ and perfect /u/ forms. In fact, this class change (containing 6 IFs) is the most important class change in the selected sample. It reflects a change that has been in progress for several centuries by which the /a/ in the past form of the class (i-a-u) is lost producing the class (i-u-u). The source of many class (i-u-u) verbs is actually the loss of the distinction between the past and perfect forms (Jespersen 1954: 49-53). Peters (2009: 26) states that most members of the class (i-u-u) are 'earlier refugees' from the class (i-a-u). This trend of losing the distinction between the past and perfect forms in English makes the class (i-u-u) much larger and more general than the class (i-a-u) (Bybee 1982).

As mentioned above, 20 instances of irregularization are attested in our sample. All these instances exhibit phonological neighbourhood effects:

---

[15] The link to this data point (http://www.mtv.com/news/2597231/star-wars-legacy-comics-star-wars-legacy-war/)
[16] The link to this data point (http://www.mywalkabout.net/2010/09/dusting-off-closed-book-chapter-6-still.html)
[17] The link to this data point (http://allaroundus.blogspot.com/2009/04/dorchester-away.html)
[18] The link to this data point (http://teenfictioncafe.blogspot.com/2010/09/boyfriends-constructing-perfect.html)

- in the past form, I get *sing-sung, ring-rung, sink-sunk, spring-sprung, shrink-shrunk, stink-stunk* in reference to *cling-clung*
- in the past form, I get *fling-flang, wring-wrang, swing-swang* in reference to *ring-rang*
- in both forms, I get *bring-brung* in reference to *cling-clung*
- in both forms, I get *cleave-cleft* in reference to *leave-left*
- in the past form, I get *beget-begat* in reference to *sit-sat*
- in both forms, I get *shit-shat, spit-spat* in reference to *sit-sat*

From connectionist perspectives, these instances of irregularization are predicted to display phonological neighbourhood effects because of their hypothesized full-form storage in the associative memory. Therefore, the activation of a word form concurrently activates all word forms that have one or more sharing properties with this word. Such instances have been attested to occur in certain connectionist simulations at rates between 3% and 24% of IVs (Rumelhart and McClelland 1986; Plunkett and Marchman 1991; Sproat 1992). Connectionist models take these high rates of irregularization as evidence for their single mechanism hypotheses, since such applications on the basis of phonological similarity evidently indicate the presence of the associative memory. These associative models do not distinguish between RVs and IVs, but pick up concurrently on patterns of various degrees of phonological similarity. Proponents of the dual mechanism approach do agree that instances of irregularization are dependent on phonological similarity, but they point out that irregularization processes are quite rare in reality (only 0.2% of IVs) in contrast with the high irregularization rates manifested in connectionist models (Xu and Pinker 1995). The findings of the analysis in this study tell us that the irregularization rate (0.08 %) is a low rate. On the basis of the evidence currently available, it seems reasonable to assume that this is accounted for under the dual mechanism approach and thus runs against connectionist approaches.

In conclusion, the results of the synchronic analyses in this section confirm that no clear relationship between word frequency and irregularization. Therefore, these results are:

- inconsistent with the assumptions of single mechanism approaches in which instances of regularization and irregularization should be at similar rates. These approaches hypothesize that RVs and IVs are processed via one single mental mechanism – either a rule-based system or an associative system. Consequently, they suggest that similar mechanisms cause the production of both regularization and irregularization. So, these verbal changes are predicted to be bidirectional; the direction of regularization and irregularization.

- consistent with the predictions of the dual mechanism approach in which verbal changes are predicted to be unidirectional. This approach hypothesizes that RVs and IVs are processed via distinct mechanisms – a rule-based system for RVs and an associative system for IVs. From the dual mechanism perspective, IVs with high frequency are less disposed to regularization processes because of their strong representations in the associative memory which make them easier to be accessed and thus more resistant to regularization processes. On the basis of the evidence obtained from the results of the synchronic analyses in sections 4.3 and 4.4, I conclude that the current movements in the English verb system are primarily unidirectional, moving towards regularization only, thus favouring the dual mechanism approach.

In the next two sections, I investigate whether or not recent (ir)regularization processes are constant over time in our selected sample. I aim to generate more evidence in favour of or against single and dual mechanism approaches in the diachronic snapshot.

4.5 The Diachronic analysis of regularization

Verbal changes are diachronically attested to be unidirectional, towards regularization only (Fries 1940; Pinker 1999; Lieberman et al. 2007; Michel et al. 2011 among other). The more infrequent the irregulars are, the more they are regularized. According to the dual mechanism approach, this is due to weak representations of these verbs in the associative memory that make them harder to be accessed and thus easier to be regularized. Lieberman et al. (2007) insist that frequency plays a clear role in regularization processes. They demonstrate that the number of IVs has gradually declined over the past centuries. In their study, they find that of the 177 IVs that existed in Old English only 98 are still irregular today. They conclude that IVs with high frequency are more prone to remain irregular over time while less frequently used ones are more prone to become regular. Also, they conclude that when frequency is accounted for, regularization rates are increasing and thus not constant over time (for more details see chapter 2).

The results of the analyses in sections 4.3 and 4.4 revealed that current verbal changes are synchronically unidirectional, towards regularization only, favouring the dual mechanism approach. The aim of this section is to investigate verb regularization in Contemporary English from the diachronic perspective. I explore whether or not recent regularization processes are constant over time in the selected sample mentioned below. So, the following question is addressed:

- In Contemporary English, are verbal changes towards regularization taking place constantly over time?

To investigate whether or not IVs are regularized regularized constantly over time, I choose the sample of this diachronic analysis from the WebCorp corpus that covers the period January 1995-December 2010. This period (16 years) is divided into two-time spans: the old span (1995-2002) and the new one (2003-2010). I select the same IVs of the sample used in the question 1 (see appendix 2). Then, in the old and new spans, word frequencies of IVs split by form and frequency are collected from the selected sample. Similarly, word frequencies of RFs split by form and frequency are collected in the old and new spans from the selected sample to draw a comparison between the two spans (see appendices 8 and 9 and for more details see chapter 3). By doing so, it can be investigated whether or not verbal changes have a constant tendency towards regularization over time. I will explore whether in the new span the regularization rate of IVs with low frequency is higher than the one in the old span. If that will be the case, there will be evidence supporting the claim of the dual mechanism approach stating that IVs with low frequency are regularized more often than IVs with high frequency as a result of retrieval failures from the associative memory.

By looking at table 25 and table 26 respectively, a general idea about frequency distributions of the verbs split by type, form and frequency in the old and new spans is obtained. The two tables demonstrate word frequencies of IVs and RFs in the two spans from the selected sample. In addition, I have calculated relative frequency of RFs, as word frequencies of RFs depend on the size of the selected sample.

Table 25: Frequency distributions of IVs and RFs in the old span from the selected sample

Type / Form	High frequency Verbs		Low frequency Verbs		Total	
	Word frequency	% of RFs	Word frequency	% of RFs	Word frequency	% of RFs
IVs	69,622		343		69,965	
RFs	215	0.31%	263	43%	478	0.68%

IVs / past	49,951		101		50,052	
RFs/ past	130	0.26%	76	43%	206	0.41%
IVs / perfect	19,671		242		19,913	
RFs/ perfect	85	0.43%	187	44%	272	1.35%

Table 26: Frequency distributions of IVs and RFs in the new span from the selected sample

Type / Form	High frequency Verbs		Low frequency Verbs		Total	
	Word frequency	% of RFs	Word frequency	% of RFs	Word frequency	% of RFs
IVs	13,133,375		56,844		13,190,219	
RFs	64,324	0.49%	49,019	46%	113,343	0.85%
IVs / past	9,310,000		15,018		9,325,018	
RFs/ past	36,914	0.39%	15,138	50%	52,052	0.56%
IVs / perfect	3,823,375		41,826		3,865,201	
RFs/ perfect	27,410	0.71%	33,881	45%	61,291	1.56%

Table 25 and table 26 display that the difference between total word frequencies of IVs in the old and new data is large. Word frequency of IVs in the old span is 69,965 that is lower than the one in the new span (13,190,219). But, this difference will not prevent us to do the statistical analysis for testing significance of the difference. A linear mixed model that will be conducted later is good to handle two different samples with two different proportions.

Let's compare regularization processes in the old span. Table 25 shows that word frequency of RFs with low frequency (263) is higher than the one with high frequency (215). Similarly, the regularization rate in low frequency group (43%) is high compared to that one in the high group (only 0.31%.). Focusing on form, the regularization rates of both forms in low frequency group are higher than the ones in high frequency group (low: 43% for the past form and 44% for the perfect form versus high: 0.26% for the past form and 0.43% for the perfect form).

Considering regularization processes in the new span, however, table 26 display that word frequency of RFs with low frequency (49,019) is lower than the one with high frequency (64,324). Yet, the regularization rate in low frequency group (46%) is higher than the one in high frequency group (only 0.49%.). Likewise, in the past and perfect forms, the regularization rates in low frequency group (50% and 45% respectively) are higher than the ones in high frequency group (0.39% and 0.71% respectively).

If we look at regularization processes in the old and new spans, there is an indication in data that there are slight increases in word frequencies of RFs with low frequency (old: 43% versus new: 46%) and with high frequency (old: 0.31% versus new: 0.49%). Thus, the regularization rate of IVs with low frequency in the new span is somewhat higher than the one in the old span. The differences in frequency distributions of RFs in both spans of our sample suggest a relationship between regularization and word frequency over time, as predicted Lieberman et al. (2007) and supporters of the dual mechanism approach: IVs with low frequency are regularized more often than IVs with high frequency. Thus, these verbal changes in the direction of regularization may be not taking place constantly over time.

A statistical model is conducted to explore the effects of time, frequency and form on relative frequencies of the verbs in the selected sample. A linear mixed model was adopted, where *relative frequency* was considered as a dependent variable and I included the variables: *time* (with two levels: old and new), *frequency* (with two levels: high and low) and *form* (with two levels: past and perfect) as fixed factors. The results of the model disclose that the main effects for *time* ($\beta = -0.01$, $t = -0.54$, $p = 0.59$) and *form* ($\beta = 0.03$, $t = 1.20$, $p = 0.23$), in addition to the effects for the interaction between them ($\beta = -0.01$, $t = -0.49$, $p = 0.62$) are not significant. Only the main effect for *frequency* ($\beta = 0.14$, $t = 4.28$, $p = 2.47e\text{-}05$) is significant. Nevertheless, the effects of its interaction with *time* and *form* ($\beta = 0.02$, $t = 0.47$, $p = 0.63$) are not significant. The lack of the effect of *time* in this model indicates that there is no impact of time on regularization processes. Hence, I conclude that the rates of verbal changes are constant between the old and new spans in the selected sample of this study. Since the interaction with *time* and *frequency is* not significant, I may also conclude that what really matters in regularization is *frequency.* This means that not only does the time has no significant effect on regularization itself, but also it does not affect the larger and significant effect of frequency on regularization. Thus, verbal changes towards regularization are constant over time in the sample of this study. These results are incompatible with results of the study of Lieberman et al. and the dual mechanism view stating that verbal changes towards regularization are not constant over time. From and the dual mechanism perspective, IVs with low frequency are increasingly regularized and moving to be more general. Nevertheless, a small difference between old span and new spans of the selected sample is reported, which goes in the direction predicted by the dual mechanism approach. This suggests that even if we take a larger sample and we find a significant effect of time on regularization, this effect will always be much smaller in magnitude than the one of frequency on the rate of regularization.

To conclude, we do not find statistical evidence that the rate of regularization in both low and high frequency verbs is not constant over time. Next, we will investigate the relationship between word frequency and irregularization over time to provide further evidence either with or against single and dual mechanism approaches.

4.6 The Diachronic analysis of irregularization

Some linguists observe that verbal changes are diachronically bidirectional, towards regularization and irregularization (Peters 2004; Nübling 2000 and Fertig 2013). For the dual mechanism perspective, verbal changes occur mostly unidirectionally over time, in that regular inflection was overapplied to IVs but not the other way round. This is due to this approach's main hypothesis that posits a fundamental distinction between regular and irregular inflections: IVs are stored in the associative memory, while RVs are generated by rules. Single mechanism approaches, however, predict that verbal changes are bidirectional. Proponents of single mechanism approaches suggest no fundamental distinction between regular and irregular inflections, and contend that both are built via one single mental mechanism – either rules or storage. Hence, these theories aim to include RVs and IVs in the past and perfect forms, as well as instances of regularization and irregularization into one single explanatory mechanism system.

The results of the diachronic analysis in section 4.4 revealed that current verbal changes in the direction of regularization are constant over time in Contemporary English, thus they speak against the dual mechanism approach. The aim of this section is to investigate verb irregularization in Contemporary English from a diachronic perspective. I explore whether or not recent irregularization processes are constant over time in the selected sample of the study mentioned below. So, the following question is addressed:

- In Contemporary English, are verbal changes towards irregularization taking place constantly over time?

To explore whether or not IVs are irregularized constantly over time, I select the sample of this diachronic analysis from the WebCorp corpus that covers the period of 16 years (1995-2010). I divide this period into two-time spans: the old span (1995-2002) and the new one (2003-2010). I select the same IVs of the sample used in the question 3 (see appendix 5). Then, in the old and new spans, I collect word frequencies of IVs split by form and frequency from the selected sample. Similarly, I collect word frequencies of IFs split by form and frequency in the old and new spans from the selected sample to draw a comparison between the two spans (see appendices 10 and 11 and for more details see chapter 3). I investigate whether or not current verbal changes have a constant trend towards irregularization over time.

In table 27 and table 28, I take an overall view about frequency distributions of the verbs split by type, form and frequency in the old and new spans. The two tables display word frequencies of IVs and IFs in the two spans from the selected sample. In addition, I have calculated relative frequencies of IFs, as word frequencies of IFs depend on the size of the selected sample.

Table 27: Frequency distributions of IVs and IFs in the old span from the selected sample

Type / Form	High frequency Verbs		Low frequency Verbs		Total	
	Word frequency	% of IFs	Word frequency	% of IFs	Word frequency	% of IFs
IVs	31,069		301		31,370	
IFs	17	0.05%	4	1.3%	21	0.07%
IVs / past	18,479		88		18,567	
IFs/ past	4	0.02%	4	4.3%	8	0.04%
IVs / perfect	12,590		213		12,803	
IFs/ perfect	13	0.10%	0	0	13	0.10%

Table 28: Frequency distributions of IVs and IFs in the new span from the selected sample

Type / Form	High frequency Verbs		Low frequency Verbs		Total	
	Word frequency	% of IFs	Word frequency	% of IFs	Word frequency	% of IFs
IVs	6,567,076		50,786		6,617,862	
IFs	4,726	0.07%	844	2%	5,570	0.08%
IVs/past	4,245,839		14,094		4,259,933	
IFs/past	1,254	0.03%	834	6%	2,088	0.03%
IVs/perfect	2,321,237		36,692		4,259,933	
IFs/perfect	3,472	0.15%	10	0.03%	2,088	0.15%

Considering irregularization processes in the old span, table 27 shows that word frequency of IFs with low frequency (4) is lower than the one with high frequency (17). However, the irregularization rate in low frequency group (1.3%) is higher than that one in the high group

(0.05%.). Focusing on form, the irregularization rates of the past form in both frequency group are equal (each 4). Whereas in the perfect form, only IFs with high frequency undergo irregularization processes (13 instances).

In the new span, similarly, table 28 displays that word frequency of IFs with low frequency (844) is lower than the one with high frequency (4,726). Yet, the irregularization rate in the low frequency group (2%) is higher than the one in the high frequency group (0.07%.). In the past form, the irregularization rate in the low frequency group (6%) is higher than the one in the high frequency group (0.03%). In the perfect form, IFs in high frequency group (0.15%) are higher than the one in low frequency group (0.03%).

Comparing irregularization processes in the old and new spans, we can see that word frequencies of IFs with low frequency (old: 1.3% versus new: 2%) and with high frequency (old: 0.05%versus new: 0.07%) are slightly different. Nevertheless, table 29 below displays that, the total irregularization rates of IFs in the old and new spans are roughly equal (old: 0.07% versus new: 0.08%). In addition, these irregularization rates are low, hence I cannot evaluate the significance of the difference between the two spans statistically. However, the results of this descriptive analysis suggest that there is no relationship between irregularization and word frequency in our selected sample: verbal changes towards irregularization are taking place constantly over time in the selected sample.

Table 29: Total frequency distributions of IVs and IFs in the new and old spans from the selected sample

	Old span		New span		Total	
	word frequency	%	word frequency	%	word frequency	%
IVs	31370	99.93%	6,617,862	99.92%	6,649,232	99.92%
IFs	21	0.07%	5,570	0.08%	5,591	0.08%

Therefore, I conclude that the results of the diachronic analyses of regularization and irregularization in sections 4.5 and 4.6 respectively are neutral, as the tendency of verbal changes towards (ir)regularization are constant over time in our sample. These results are incompatible with predictions of the dual mechanism approach that confirm unidirectionality of language change; towards regularization only. In addition, they speak against single mechanism approaches that confirm bidirectionality of language change in the direction of regularization and irregularization. Nevertheless, in the next steps, I can only provide a further descriptive analysis of irregularization processes in the two spans.

To gain more in-depth understanding of irregularization processes over time, I look at all possible changes of irreguarization in the selected verbs classified into the 35 classes in each span (see table 10 and for more details see chapter 3). I check whether the changed classes of IVs display any phonological neighbourhood effects in each span because of being stored in the associative memory as predicted by dual and connectionist models. Table 30 exhibits which of the 35 classes are changed and which do not display word frequencies of IFs within the changed classes in both spans. Of the 35 classes, 18 show no class change, while 6 display various types of irregularization (see appendices 10 and 11). The changed classes with their word frequencies in both spans are:

- 777 instances change their classes 1A-1 (like, *put-put-put*) into 2C-2 (like, *sit-sat-sat*)
- 549 instances change their classes from 2C-1 (like, *cling-clung-clung*) into 3A-2 (like, *ring-rang-rung*)
- 32 instances change their classes 2C-9 (like, *buy-bought-bought*) into 2C-1 (like, *cling-clung-clung*)
- 15 instances change their classes 2D-4 (like, *speak- spoke-spoken*) into 2C-6 (like, *meet-met-met*)
- 2,937 instances change their classes 2D-5 (like, *get-got-got*) into 2C-2 (like, *sit-sat-sat*)
- 1,269 instances change their classes from 3A-2 (like, *ring-rung-rung*) into 2C-1 (like, *cling-clung-clung*)

Table 30: The changed and unchanged classes of IVs in both spans from the selected sample

N.	Class	Ex.	Appearance	Class-change	Ex.	Word freq. of IFs
1	1A-1	put-put-put	Yes	2C-2	sit-sat-sat	777
2	1B-1	beat-beat-beaten	No	-	-	-
3	1C-1	learn-learned-learned	No	-	-	-
4	1C-2	spoil-spoilt-spoilt	No	-	-	-

5	1C-3	clap-clapt-clapt	No	-	-	-
6	1C-4	bless-blest-blest	No	-	-	-
7	1C-5	make-made-made	No	-	-	-
8	1C-6	spend-spent-spent	No	-	-	-
9	1C-7	pay-paid-paid	No	-	-	-
10	2A -1	come-came-come	No	-	-	-
11	2A -2	run-ran-run	No	-	-	-
12	2B -1	take-took-taken	No	-	-	-
13	2B -2	give-gave-given	No	-	-	-
14	2B -3	know-knew-known	No	-	-	-
15	2B -4	draw-drew-drawn	No	-	-	-
16	2B -5	see-saw-seen	No	-	-	-
17	2B -6	eat-ate-eaten	No	-	-	-
18	2B -7	fall-fell-fallen	No	-	-	-
19	2C-1	cling-clung-clung	Yes	3A-2	ring-rang-rung	549
20	2C-2	sit-sat-sat	No	-	-	-
21	2C-3	stand-stood-stood	No	-	-	-
22	2C-4	hear-heard-heard	No	-	-	-
23	2C-5	find-found-found	No	-	-	-
24	2C-6	meet-met-met	No	-	-	-
25	2C-7	lose-lost-lost	No	-	-	-
26	2C-8	tell-told-told	No	-	-	-
27	2C-9	buy-bought-bought	Yes	2C-1	cling-clung-clung	32
28	2D-1	wear-wore-worn	No	-	-	-
29	2D-2	hide-hid-hidden	Yes	2D-2	light-lit-lit	25
30	2D-3	lie-lay-lain	No	-	-	-
31	2D-4	speak-spoke-spoken	Yes	2C-6	meet-met-met	15
32	2D-5	get-got-got	Yes	2C-2	sit-sat-sat	2,937
33	3A-1	write-wrote-written	No	-	-	-
34	3A-2	ring-rang-rung	Yes	2C-1	cling-clung-clung	1,270
35	3A-3	fly-flew-flown	No	-	-	-
Total word frequency of IFs						5616

Within the 6 changed classes, only 18 IVs that undergo various types of irregularization in both forms namely *shit, spit, spit, fling, wring, swing, strike, bring, bitten, cleave, forget, beget, sing, ring, sink, spring, shrink* and *stink.* In the old span, only 6 IFs are attested (total occurrences: 21), whereas in the new span 20 IFs are found (total occurrences: 5,595) (see appendices 10 and 11). Table 31 shows word frequencies of IVs and IFs the old and new spans to help us draw comparisons. In addition, the direction of vowel change in the changed class is mentioned; whether it is towards increasing (like *spit-spit-spit* (one-vowel paradigm) is changed into *spit-spat-spat* (two-vowel paradigm)) or decreasing (like sing-sang-sung (three-vowel paradigm) is changed into sing-sung-sung (two-vowel paradigm)).

Table 31: Word frequencies of IVs and IFs in the old and new spans from the selected sample

Class	IVs	Form	Word freq. old span	Word freq. new span	Class-change	IFs	Word freq. old span	Word freq. new span	Direction
One-vowel paradigm									
1A-1	shit	past	9	2137	2C-2	shat	3	237	increase
	spit	past	4	1017		spat	1	394	increase
	spit	perfect	1	317		spat	-	142	increase
Two-vowel paradigm									
2C-1	flung	past	2	305	3A-2	flang	-	2	increase
	wrung	past	0	50		wrang	-	2	increase
	swung	past	3	1672		swang	-	1	increase
	struck	perfect	28	5496	3A-1	stricken	5	539	increase
2C-9	brought	past	200	40093	2C-1	brung	-	27	-
	brought	perfect	126	25818		brung	-	5	-
2D-2	bitten	perfect	5	1257	2D-2	bit	-	25	-
2D-4	clove	past	0	76	2C-6	cleft	-	5	-
	cloven	perfect	0	1		cleft	-	10	-
2D-5	forgotten	perfect	76	20773	2C-7	forgot	8	2786	-
	begot	past	79	15596	2C-2	begat	3	151	-
Three-vowel paradigm									
3A-2	sang	past	24	5457	2C-1	sung	-	3	decrease
	rang	past	5	2269		rung	-	2	decrease
	sank	past	4	1204		sunk	-	587	decrease
	sprang	past	7	596		sprung	-	3	decrease
	shrank	past	1	268		shrunk	-	4	decrease
	stank	past	1	169		stunk	1	670	decrease
Total word frequency			575	12457 1			21	5,595	

Table 31 displays that in the old and new spans, I get:

- 3 verbs change their classes from 1A-1 into 2C-2 (old: 4 instances versus new: 773 instances)
- 3 verbs change their classes from 2C-1 into 3A-2 (only new: 5 instances)
- 1 verbs change its class from 2C-1 into 3A-1 (old: 5 instances versus new: 539 instances)

- 2 verbs change their classes from 2C-9 into 2C-1 (only new: 32 instances)
- 2 verbs change their classes from 2D-4 into 2C-6 (only new:15 instances)
- 2 verbs change their classes from 2D-5 into 2C-2 and the other into 2C-7 (old: 11 instances versus new: 2,937 instances)
- 6 verbs change their classes from 3A-2 into 2C-1 (old: 1 instances versus new 1,269 instances)
- 1 verb goes under simplification within the same class (2D-2) (from *bite-bit-bitten* to *bite-bit-bit*) (new: 25 instances)

In both spans, all instances of irregularization attested in our sample display phonological neighbourhood effects:

- In the old and new spans, I get *shit-shat, spit-spat* in reference to *sit-sat.*
- Only in the new span, I get fling-flang, wring-wrang, swing-swang in reference to ring-rang
- Only in the new span, I get *bring-brung* in reference to *cling-clung*
- Only in the new span, I get *cleave-cleft* in reference to *leave-left*
- In the old and new spans, I get *forget-forgot* in reference to *get-got*
- In the old and new spans, I get *beget-begat* in reference to *sit-sat*
- In the old and new spans, in the past form, I get *sing-sung, ring-rung, sink-sunk, spring-sprung, shrink-shrunk, stink-stunk* in reference to *cling-clung*

From dual and connectionist perspectives, these instances of irregularization are predicted to display phonological neighbourhood effects, as they are stored in the associative memory in which the activation of a word form can simultaneously activate all word forms that share one or more of the phonological properties of the word. However, followers of the dual mechanism approach, unlike connectionism, predict that instances of irregularization are quite rare in reality (Xu and Pinker 1995). The results of this diachronic analysis display that the low rates of irregularization in both spans (old: 0.07% versus new: 0.08%) are accounted for under the dual mechanism approach and thus run against connectionist approaches.

Focusing on the direction of vowel change of IFs, the following verbs are changed from:

- three-vowel paradigm to two-vowel paradigm, 6 IFs namely *sing, ring, sink, spring, shrink, stink*

- one-vowel paradigm to two-vowel paradigm, 3 IFs namely *shit, spit, spit*
- two-vowel paradigm to three-vowel paradigm, 4 IFs namely *fling wring swing strike*
- remain with the same two-vowel paradigm, 5 IFs namely *bring cleave forget beget, bit*

As the changes seem to be almost equal in both direction (7 towards increasing direction and 6 towards decreasing direction, we can say that there is no clear direction of vowel change in irregularization processes.

Altogether, in the synchronic snapshot, in sections 4.3 and 4.4, (ir)regularization processes have been explored to check whether there a relationship between word frequency and (ir)regularization processes in Contemporary English. The results of the synchronic analysis in section 4.3 disclose that there is a relationship between word frequency and regularization processes with a rate of 0.98% of IVs. Hence, IVs with low frequency are generally regularized more often than IVs with high frequency in Contemporary English. However, the results of the synchronic analysis in section 4.4 confirm that there is no clear relationship between high word frequency and irregularization in the selected sample. The irregularization rate (0.08%) is very low in our sample. Therefore, I conclude that verbal changes are synchronically unidirectional in Contemporary English, in the direction of regularization. These results are consistent with the predictions of the dual mechanism approach speaking against single mechanism approaches. The dual mechanism approach hypothesizes that RVs and IVs are processed via distinct mechanisms: a rule-based system for RVs and an associative system for IVs. Single mechanism approaches, nevertheless, include both RVs and IVs into one single explanatory mechanism system: either rules or storage.

In the diachronic snapshot, in sections 4.5 and 4.6, (ir)regularization processes have been investigated to test whether there a relationship between word frequency and (ir)regularization processes over time in Contemporary English: *are verbal changes towards (ir)regularization occurring constantly over time in Contemporary English?* The results of these analyses disclose that verbal changes towards regularization and irregularization are constant over time; they are neutral. The results of the diachronic analyses are incompatible with predictions of both the dual mechanism approach (unidirectionality of language change) and single mechanism approaches (bidirectionality of language change). Nevertheless, within each span, it seems that the low frequent irregulars are (ir)regularized more often than the high frequent ones in our sample. In addition, certain classes of IVs that are changed into IVs of other classes seem to be phonologically related to some extent. More specifically, 2C-1 displays a trend towards 3A-2 and vice versa.

5 Chapter Five: Conclusion and suggestions for further research

The primary goal of the current study is to empirically evaluate the debate on the dual versus the single mechanism approaches in the human mind (Chomsky and Halle 1968; Halle and Mohanan 1985; Rumelhart and McClelland 1988; Pinker and Prince 1988; Pinker 1999): *Does the human language system exploit two cognitive mechanisms (rules and storage) or a single mechanism (either rules or storage)?*

5.1 5.1 Conclusion

In an attempt to answer this question, the current study has focused on language change (and therefore language acquisition). I have tracked synchronic developments and diachronic movements in the English verb system currently. My overall goal was to check whether verbal changes occur uni-directionally (towards regularization) or bi-directionally (towards both regularization and irregularization). To this end, this study has sought to find out whether RVs versus IVs display frequency effects (which is taken as a reflection of storage) or not (which is taken as a reflection of computation). The presence of frequency effects, therefore, can be considered as a diagnostic of the storage and retrieval of these forms from the associative memory, whereas the absence of frequency effects indicates their composition.

Different approaches have been suggested to account for processing differences between inflectional types. These approaches differ in frequency effects they predict for inflectional types, as summarized in table 32 below. According to connectionist approaches, all inflected forms are processed in the associative memory (Rumelhart and McClelland 1986; Plunkett and Marchman 1993; Elman 1999). Hence, both IVs and RVs are predicted to display frequency effects (as an outcome of storage). On the other hand, rule-based approaches predict that all inflected forms are generated by rules and accordingly IVs and RVs are predicted to exhibit no frequency effects (as an outcome of computation). Thus, these two approaches propose that the same mechanisms underlie the production of regularization and irregularization processes. Accordingly, IVs are predicted to be regularized at the same rate as they are irregularized. This bidirectional prediction by single system approaches is in conflict with the prediction by the dual system approach that verbal changes mostly occur unidirectionally. Along with the dual approach, two routes are proposed: one route is rule-governed that enables the formation of RVs, whereas the second route relates to the memory system of IVs. IVs can be produced correctly if they are memorized and retrieved successfully before the rule-governed route creates forms of regularization. Thus, IVs with low frequency are predicted to be regularized more often than the ones with high frequency. However, instances of irregularization are predicted to be exceedingly rare (no more than 0.2%; Xu and Pinker 1995). Therefore, the dual approach predicts that the regularization rate should be higher than the irregularization one. The different hypotheses of single-dual system approaches can be formulated in the following table:

Table 32: The different predictions of single-dual mechanism approaches regarding the production of RVs versus IVs and RFs versus IFs

Approach	Rules only	Storage only	Rules & Storage
Type	RVs and IVs	RVs and IVs	rules for RVs storage for IVs
Frequency effects	no	yes	only for IVs and RFs
Phonological similarity	no	yes	only for IVs and IFs
(Ir)regularization rates	similar	similar	RFs is higher than IFs
Direction of verbal changes	bidirectional (RFs and IFs)	bidirectional (RFs and IFs)	unidirectional (only RFs)

It is well-known that single-dual mechanism approaches have originally been set up as models of L1 processing. The assumptions of these approaches are generally proposed to have universal features of human language. Hence, these assumptions are meant to carry universal legitimacy and accordingly they should hold for language processing of L2 and multilingual learners as well. One can assume that L2 and multilingual learners may already employ these mechanisms of language processing in their native language. Recently, the advocates of the dual mechanism account have extended their assumptions of the two distinct mechanisms of morphological processing in L2 acquisition. For example, Clahsen and Felser (2006) claim that morphological processing in L1 and L2 are similar and display a dissociation of rule-based and associative patterning. This study is a contribution to single-dual mechanism debate of language processing by investigating the possibility of a relationship between word frequency and (ir)regularization in the English verbal system in a multilingual environment. It demonstrates how language acquisition research in multilingual environment forms as an additional testing ground to evaluate single and dual mechanism approaches. For this purpose, I ran a corpus study on the synchronic and diachronic levels in the internet space where multilingualism is diffused (Cenoz et al. 2003; Aronin and Singleton 2008; Auer and Wei 2007;

Cook 1992; Grosjean 1982, 2010) and accordingly I assume that verbal changes are predicted to occur more than usual (see chapters 1 and 3 for more details).

On the synchronic level of this corpus study, the following questions have been addressed:

1. Are IVs generally more frequent than RVs in the past and perfect forms in Contemporary English?

The results of this research question have demonstrated that there is a relationship between word frequency and irregularity. IVs have displayed frequency effects, whereas RVs do not (mean frequency of IVs 58,913 versus mean frequency of RVs 9,958). Hence, IVs are generally more frequent than RVs in the selected sample. This result is additionally supported by the fact that the ten most common verbs (*be, have, say, do, get, make, go, think, come* and *take*) are all irregular in the selected sample. These top ten IVs comprise 71% of the total word frequencies of 30 verbs with the highest word frequency in the past form in the 500-verb sample of this study (see chapter 4 for more details). Yet, the past and perfect forms display no frequency effects on (ir)regularity in this study.

2. Do regularization processes take place in Contemporary English? If so, are IVs with low frequency regularized more often than IVs with high frequency in the past and perfect forms?

A relationship between word frequency and regularization processes has been attested in this corpus study. Generally, IVs with low frequency (60%) are more prone to be regularized than the ones with high frequency (0.43%). Again, no frequency effects of the past and perfect forms are on (ir)regularity.

3. Do regularization processes occur more frequently in the cases where IVs and RFs show no vowel change in Contemporary English?

Taking a deeper step into the nature of regularization processes, I investigate the impact of vowel change on the retention of regularization processes in the 42 doublet verbs that can be both regular and irregular in the English language like *burned/burnt* and *lighted/lit*. The results of the descriptive analysis of the selected data do suggest that the verbs with no vowel change (like *burned/burnt*) are less resistant to regularization processes and hence are regularized more often than those with vowel change (like *lighted/lit*) in both forms. These results are consistent with results of the study of De Clerck and Vanopstal (2015) in which a relationship between the salience of vowel change and regularization processes and is attested (see chapter 2 for more details).

4. Do irregularization processes take place in Contemporary English? If so, are IVs with low frequency regularized more often than IVs with high frequency in the past and perfect forms?

In this corpus study, no relationship between word frequency and irregularization processes in the past and perfect forms has been attested. In the selected sample of this study, there are few instances of irregularization; only 0.08 % of IVs). Nevertheless, I have explored whether the instances of irregularization are closely analogized from existing irregular classes: *Do instances of irregularization display any phonological neighbourhood effects?* In the selected sample, all the attested instances of irregularization display phonological neighbourhood effects. Both followers of dual and connectionist models assume that these instances should exhibit phonological neighbourhood effects because IVs are stored in the associative memory and accordingly the activation of one word can activate all other words that share one or more of the phonological properties of the word. In this respect, Xu and Pinker (1995) claim that:

> *Irregular forms are stored as memorized linked pairs of lexical entries in the mental dictionary. The patterns shown across the irregulars are due to the associative nature of memory: when X is linked to Y, the properties of X are also linked to the properties of Y, so that new items similar to X (that is, sharing properties with X) have some probability of activating the properties of Y* (Xu and Pinker 1995: 553).

Nevertheless, from the dual mechanism perspective, unlike under connectionism, it is predicted that instances of irregularization are quite rare. As mentioned previously, in the study of Xu and Pinker (1995) the irregularization rate is 0.2% and this low rate contrasts with the high irregularization rates presented in the studies of connectionist models (between 3% and 24% of IVs in Rumelhart and McClelland 1986; Plunkett and Marchman 1991; Sproat 1992). In this study, the irregularization rate is also relatively low (only 0.08 % of IVs) and this low rate is accounted for under the dual mechanism approach and thus runs against connectionist approaches.

Additionally, of the 35 classes of IVs (see table 10), 18 display no class change, whereas 6 display different kinds of irregularization: 1A-1 (*put-put-put*) into 2C-2 (*sit-sat-sat*), 2C-1 (*cling-clung-clung*) into 3A-2 (*ring-rang-rung*), 2C-9 (*buy-bought-bought*) into 2C-1 (*cling-clung-clung*), 2D-4 (*speak- spoke-spoken*) into 2C-6 (*meet-met-met*), 2D-5 (*get-got-got*) into 2C-2 (*sit-sat-sat*) and 3A-2 (*ring-rung-rung*) into 2C-1 (*cling-clung-clung*). All the instances of changed classes exhibit phonological neighbourhood effects. For instance, in the past form, I obtain *sing-sung, ring-rung, sink-sunk, spring-sprung, shrink-shrunk, stink-stunk* in reference to *cling-clung.* Of the 6 changed classes, the most important classes of IVs that are changed into IVs of other classes are 3A-2 (i-a-u like *ring-rang-rung*) into 2C-1 (i-u-u like *cling-clung-clung*) and vice versa. For example, the verb *stink* shows the strongest trend of class change from the three-vowel paradigm (*stink-stank-stunk*) into the direction of the two-vowel paradigm (*stink-stunk-stunk*). This verb gets the highest percentage (79.8% of IVs) compared to other data points of irregularization.

In the light of the obtained results of this synchronic study, I conclude that indeed there is a relationship between word frequency and regularization processes. In the sample, verbal changes are unidirectional, moving in the direction of regularization. These findings can be evaluated as unsupportive of single system approaches that, as mentioned at various points in the preceding discussions, basically theorize that RVs and IVs are processed via the same mechanism – either storage or rules. Advocates of these models predict that all inflected forms should either show frequency effects (suggesting storage) or not (suggesting rules). Hence, these verbal changes are predicted to be bidirectional; in both directions. However, under the dual-mechanism view, these findings are predicted because what would be supposed is a dissociation between RVs and IVs that manifests itself in frequency effects for IVs and no frequency effects for RVs. Therefore, the synchronic data of this study suggest that current English verb system are unidirectional in Contemporary English, moving towards regularization only.

On the diachronic level of the current study, the following questions have been formulated:

5. Are verbal changes towards regularization taking place constantly over time in Contemporary English?
6. Are verbal changes towards irregularization taking place constantly over time in Contemporary English?

The results of the diachronic analysis of regularization in this study have revealed that the rates of RFs with low frequency (old span: 43% versus new span: 46%) and with high frequency (old span: 0.31% versus new span: 0.49%) are slightly increasing over time. These differences are not statistically significant. This tells us that verbal changes towards regularization are constant over time in our sample and thus speaks against the assumption of the dual mechanism approach: IVs with low frequency are predicted to be regularized more often than IVs with high frequency. From the dual mechanism perspective, therefore, verbal changes towards regularization are predicted to not be constant over time. Likewise, the results of the diachronic analysis of irregularization demonstrate that IFs with low frequency (old span: 1.3% versus new span: 2%) and with high frequency (old span: 0.05%versus new span: 0.07%) are roughly similar and hence tend to be constant over time. These differences cannot be evaluated statistically, since they are low. The findings of the diachronic analyses of (ir)regularization are rather unpredicted, since verbal changes are predicted to be either unidirectional (favouring the dual mechanism approach) or bidirectional (favouring single mechanism approaches). However, a possible explanation for these unpredicted results may be related to the fact that the time frame (only 16 years) selected from WebCorp (the corpus of this study) was rather small (see chapter 3 for more details). Therefore, in such a short time frame, it is not sufficient to arrive at powerful conclusions. For a future investigation, it is definitely necessary to employ a larger time frame.

Taken together, the general conclusion of the present study is that the results obtained from the synchronic study of exploring frequency effects of (ir)regularity and (ir)regularization in English verbal system are fully compatible with the dual-mechanism approach. This approach predicts dissociations between fully stored IVs in the associative memory and RVs that are theorized being rule-governed. Yet, the findings of the diachronic analysis of the study present an inconclusive picture. The data of (ir)regularization processes are constant over time and hence are not supportive of single-dual mechanism approaches.

Generally speaking, in this study, the results of (ir)regularization (as a testing ground for single-dual mechanism approaches of language processing) point to the likelihood that the dual mechanism approach of language processing is in many ways consistent and can account for any linguistic phenomena not only in monolingual environments, but also multilingual ones.

5.2 Suggestions for further research

In the current study, I have used the linguistic phenomenon of the English past and perfect forms in the verb system to investigate morphological processing in multilingual space. Another area that will be potentially fruitful to investigate morphological processing is the English nominal system in the same space. This is due to the fact that the distribution of the morphological structure involved in this system has certain similarities to the English past and perfect forms.

Within the framework of the dual mechanism approach, the predictions of the past tense should principally be the same for nominal inflection. Take for instance number inflection. Regular plurals are computed by the default rule (add *–s* to the noun stem e.g., *table-tables, door-doors, box-boxes*), while irregular plural nouns are assumed to be stored in the associative memory (e.g., *child-children, ox-oxen, tooth-teeth*). Instances of regularization in the noun system, like in the verb system, are hypothesized to be produced by adding the default suffix to irregular noun stems whenever there is failure retrieval of the appropriate irregular form from the associative memory. Marcus (1995) shows that the regularization rates in the English noun system are almost identical to the regularization rates in the English verb system (8.5%29 versus 7.3%, respectively). These results are supportive of the dual mechanist view. This is because type and word frequencies of English plural regular nouns are higher than type and word frequencies of English plural irregular nouns (in contrast to the English past tense system, in which irregular type and word frequencies are higher). The rates of these frequencies in Marcus's (1995: 449) study are summarized in the following table:

Table 33: The rates of type and word frequencies of English noun and verb systems (Based on Marcus's (1995: 449) study)

System	Type	Type frequency	word frequency
English noun system	Regular	98%	97%
	Irregular	2%	3%
English verb system	Regular	86%	40%
	Irregular	14%	60%

For the dual mechanism perspective, since regulars are produced by means of rules that are theoretically frequency-insensitive. Therefore, the regularization rates that are not affected by the differences of type and word frequencies between the English noun system and the English verb system are in line with assumptions of the dual mechanism approach. Nevertheless, from connectionist perspectives, this must result in a higher regularization rate in the English noun system. Supporters of connectionism predict that higher type and word frequencies of English plural nouns should lead to stronger connections of these forms in the associative memory. Consequently, these strong connections may cause constant generalizations to new instances. Hence, the English noun system is another good ground to test these predictions and accordingly to collect more evidence for/against approaches of morphological processing in the multilingual environment.

References

Aalberse, S. (2010). Inflectional economy and politeness : morphology-internal and morphology-external factors in the loss of second person marking in Dutch. University of Amsterdam.

Aitchison, J. (2004). *Language change: Progress or decay?* (Third edit). Cambridge, United Kingdom: Cambridge University Press.

Albright, A., & Hayes, B. (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition*, *90*(2), 119–161.

Andersen, H. (1973). Abductive and Deductive Change. *Language*, *49*(4), 765. https://doi.org/10.2307/412063

Aronin, L., & Singleton, D. (2008). The complexity of multilingual contact and language use in times of globalization. *Conversarii. Studi Linguistici*, *2*, 33–47.

Auer, P., & Wei, L. (2007). *Handbook of Multilingualism and Multilingual Communication* (Vol. 5). Walter de Gruyter.

Beck, M.-L. (1997). Regular verbs, past tense and frequency: tracking down a potential source of NS/NNS competence differences. *Second Language Research*. Sage Publications, Ltd. https://doi.org/10.2307/43104536

Berko, J. (1958). The child's learning of English morphology. *Word*, *14*(2–3), 150–177.

Blom, E., Polišenská, A. D., Weerman, A. F., Blom, E., Polišenská, D., & Weerman, F. (2007). Effects of age on the acquisition of agreement inflection. *Morphology*, *16*, 313–336. https://doi.org/10.1007/s11525-007-9110-1

Bock, K., Eberhard, K. M., Cutting, J. C., Meyer, A. S., Schriefers, H., Dell, G., … Van, M. (2001). Some Attractions of Verb Agreement. *Cognitive Psychology*, *43*, 83–128. https://doi.org/10.1006

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, *10*(5), 425–455.

Bybee, J. L., & Hopper, P. J. (2001). *Frequency and the emergence of linguistic structure* (Vol. 45). John Benjamins Publishing.

Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science*, *27*(6), 843–873.

Carroll, R., Svare, R., & Salmons, J. C. (2012). Quantifying the evolutionary dynamics of German verbs. *Journal of Historical Linguistics*, *2*(2), 153–172. https://doi.org/10.1075/jhl.2.2.01car

Chomsky, C. (1970). Reading, writing, and phonology. *Harvard Educational Review*, *40*(2), 287–309.

Chomsky, N. (1981). Lectures on binding and government. *Dordrecht: Foris*.

Chomsky, N. (1986). Some observations on language and language learning: Reply to Macnamara, Arbib, and Moore and Furrow. *New Ideas in Psychology*, *4*(3), 363–377. https://doi.org/10.1016/0732-118X(86)90046-2

Chomsky, N. (1995). *The Minimalist Program*. Cambridge: MIT Press. https://doi.org/10.2307/415885

Chomsky, N. (2000). Minimalist inquiries: The framework. Step by step: Essays on minimalist syntax in honor of Howard Lasnik, eds. R. Martin, D. Michaels & J. Uriagereka, 89–155. Cambridge, MA: MIT Press.

Chomsky, N., & Collins, C. (2001). *Beyond explanatory adequacy* (Vol. 20). MITWPL.

Chomsky, N., & Halle, M. (1968). The sound pattern of English.

Clahsen, H. (1995). German plurals in adult second language development. In *The Current State of Interlanguage* (p. 123). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/z.73.11cla

Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences*, *22*(6), 991–1013. Retrieved from https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/lexical-entries-and-rules-of-language-a-multidisciplinary-study-of-german-inflection/6608FA17C363432CF92BCE3322541897

Cook, V. J. (1992). Evidence for Multicompetence. *Language Learning*, *42*(4), 557–591. https://doi.org/10.1111/j.1467-1770.1992.tb01044.x

Corbett, G. G. (2006). *Agreement* (Vol. 109). Cambridge University Press.

Croft, W. (2000). *Explaining language change: An evolutionary approach*. Pearson Education.

Crystal, D. (2003). *English as a global language* (Second). Cmabridge: CAMBRIDGE UNIVERSITY PRESS. Retrieved from https://s3.amazonaws.com/academia.edu.documents/44242470/david_crystal.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1515951148&Signature=0NbPiC9wNyd1eGyA%2F9QiwxZa3oE%3D&response-content-disposition=inline%3Bfilename%3DEnglish_as_a_global_language_S

Crystal, D. (2004). *Language and the Internet*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511487002

Crystal, D. (2011). *Internet linguistics : a student guide*. Routledge. Retrieved from https://www.routledge.com/Internet-Linguistics-A-Student-Guide/Crystal/p/book/9780415602716

De Clerck, B., & Vanopstal, K. (2015). Patterns of regularisation in British, American and Indian English. *Grammatical Change in English World-Wide*, *67*, 335.

Di Sciullo, A.-M., & Williams, E. (1987). *On the definition of word* (Vol. 14). Springer.

Don, Jan; Drijkoningen, Frank; Zonneveld, W. (1994). *Inleiding in de generatieve morfologie*. Bussum : Coutinho, 1994. Retrieved from http://www.cageweb.be/catalog/cgw01:000560132

Embick, D., & Halle, M. (2005). On the status of stems in morphological theory. *AMSTERDAM STUDIES IN THE THEORY AND HISTORY OF LINGUISTIC SCIENCE SERIES 4*, *270*, 37.

Embick, D., & Marantz, A. (2005). Cognitive neuroscience and the English past tense: Comments on the paper by Ullman et al. *Brain and Language*, *93*(2), 243–247.

Embick, D., & Noyer, R. (2007). Distributed morphology and the syntax/morphology interface. In *The Oxford handbook of linguistic interfaces* (pp. 289–324). Oxford University Press Oxford. Retrieved from https://pdfs.semanticscholar.org/d423/f3d93530f4520589fbeb00304a1d5b49ad30.pdf

Fertig, D. (2013). *Analogy and morphological change*. Edinburgh University Press. Retrieved from http://hosting03.snu.ac.kr/~korean/old/data/morphology/Fertig(2013)Analogy_and_Morphological_Change.pdf

Fleischhauer, E. (2013). Morphological processing in children: an experimental study of German past participles. University of Potsdam. Retrieved from https://publishup.uni-potsdam.de/files/6807/fleischhauer_diss.pdf

Grosjean, F. (1982). *Life with two languages: An introduction to bilingualism*. Harvard University Press.

Grosjean, F. (2010). *Bilingual: Life and reality*. Cambridge: Harvard University Press.

Halle, M. (1962). Phonology in Generative Grammar. *WORD*, *18*(1–3), 54–72. https://doi.org/10.1080/00437956.1962.11659765

Halle, M., & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale & S. J. Keyser (Eds.), *The view from building 20* (pp. 111–176). Cambridge, MA: The MIT Press. Retrieved from https://nyuscholars.nyu.edu/en/publications/distributed-morphology-and-the-pieces-of-inflection

Halle, M., & Marantz, A. (1994). Some key features of Distributed Morphology. *MIT Working Papers in Linguistics*, *21*(275), 88.

Halle, M., & Mohanan, K. P. (1985). Segmental phonology of modern English. *Linguistic Inquiry*, *16*(1), 57–116.

Hammarberg, B. (2010). The languages of the multilingual: Some conceptual and terminological issues. *IRAL-International Review of Applied Linguistics in Language Teaching*, *48*(2–3), 91–104.

Harley, H. (2010). *The syntax/morphology interface*. University of Arizona.

Harley, H., & Noyer, R. (1999). Distributed morphology. *Glot International*, *4*(4), 3–9.

Hock, H. H. (1991). *Principles of historical linguistics*. Walter de Gruyter.

Hyltenstam, K., & Abrahamsson, N. (2003). Maturational Constraints in SLA. In *The Handbook of Second Language Acquisition* (pp. 538–588). Oxford, UK: Blackwell Publishing Ltd. https://doi.org/10.1002/9780470756492.ch17

Jakobson, R. (1929). *Remarques aur l'evolution phonologique de Russe comparee a celle des autres langues slaves.* na.

Jespersen, O. (1965). *A Modern English Grammar - On Historical Principles - Part VI - Morphology*. London: George Allen and Unwin Ltd.

Joanisse, M. F., & Seidenberg, M. S. (1999). Impairments in verb morphology after brain injury: a connectionist model. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(13), 7592–7. https://doi.org/10.1073/PNAS.96.13.7592

Johnson, J. S., & Newport, E. L. (1991). Critical period effects on universal properties of language: The status of subjacency in the acquisition of a second language. *Cognition*, *39*(3), 215–258. https://doi.org/10.1016/0010-0277(91)90054-8

Kiparsky, P. (1982). Word-formation and the lexicon. Mid-America Linguistics Conference. Retrieved from https://kuscholarworks.ku.edu/bitstream/handle/1808/20827/MALC_1982_3-29_Kiparsky.pdf?sequence=1

Kremers, J. (2003). The noun phrase in Arabic: a minimalist approach. *Doctoral Dissertation, University of Nijmegen*.

Kroch, A., & Taylor, A. (1997). *Verb movement in Old and Middle English: Dialect variation and language contact*. Cambrige: Cambridge University Press.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Dartmouth Publishing Group.

Labov, W. (1994). Principles of language change: Internal factors. Oxford: Blackwell.

Lapointe, S. (1981). The representation of inflectional morphology within the lexicon. In *North East Linguistic Conference* (Vol. 11, pp. 190–204).

Lapointe, S. G. (1980). *A theory of Grammatical Agreement*. University of Massachusettsm Amherst.

Lieber, R. (1992). *Deconstructing morphology : word formation in syntactic theory*. Chicago: University of Chicago Press.

Lieberman, E., Michel, J.-B., Jackson, J., Tang, T., & Nowak, M. A. (2007). Quantifying the evolutionary dynamics of language., *449*(7163), 713–6. Retrieved from http://www.jbe-platform.com/content/journals/10.1075/jhl.2.2.01car

Lightfoot, D. (1999). *The development of language: Acquisition, change, and evolution*. Wiley-Blackwell.

Lightfoot, D. W. (1979). Principles of diachronic syntax. *Cambridge Studies in Linguistics London*, *23*.

Longobardi, G. (1999). Syntactic change: The inertial theory. In *21st annual meeting of the DGfS, Workshop on Principles of Syntactic Reconstruction*.

Longobardi, G. (2001). Formal Syntax, Diachronic Minimalism, and Etymology: The History of French, Chez. *Linguistic Inquiry*, *32*(2), 275–302. https://doi.org/10.1162/00243890152001771

Lu, H., Chen, D., & Holyoak, K. J. (2012). Bayesian analogy with relational transformations. *Psychological Review*, *119*(3), 617–648. https://doi.org/10.1037/a0028719

MacWhinney, B. (2000). The CHILDES project: Tools for analyzing talk: Volume I: Transcription format and programs, volume II: The database. MIT Press.

MacWhinney, B., & Leinbach, J. (1991). Implementations are not conceptualizations: Revising the verb learning model. *Cognition*, *40*(1–2), 121–157. https://doi.org/10.1016/0010-0277(91)90048-9

Marcus, G. F. (1995). The acquisition of the English past tense in children and multilayered connectionist networks. *Cognition*, *56*(3), 271–279.

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., Xu, F., & Clahsen, H. (1992). Overregularization in Language Acquisition. *Monographs of the Society for Research in Child Development*, *57*(4), i-178. https://doi.org/10.2307/1166115

McClelland, J. L., & Patterson, K. (2002). Rules or connections in past-tense inflections: what does the evidence rule out? *Trends in Cognitive Sciences*, *6*(11), 465–472. https://doi.org/10.1016/S1364-6613(02)01993-9

Michael, L. (2014). *Social dimensions of language change*. (C. Bowern & B. Evans, Eds.), *Routledge Handbook of Historical Linguistics*. Routledge.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., … Orwant, J. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*(6014), 176–182.

Milroy, J., & Milroy, L. (1985). *Authority in language: Investigating language prescription and standardisation*. Routledge.

Moscoso del Prado Martín, F., Bertram, R., Häikiö, T., Schreuder, R., & Baayen, R. H. (2004). Morphological Family Size in a Morphologically Rich Language: The Case of Finnish Compared With Dutch and Hebrew. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1271–1278. https://doi.org/10.1037/0278-7393.30.6.1271

Müller, F. M. (1890). *Three Lectures on the Science of Language and Its Place in General Education: Delivered at the Oxford University Extension Meeting 1889, with a Supplement" My Predecessors", an Essay on the Genesis of the Idea of the Identity of Thought and Language in th* (Vol. 54). Chicago: The Open Court Publishing Company.

Orbjörg Hróarsdóttir, ! (2009). Notes on language change and grammar change. *Working Papers in Scandinavian Syntax*, *83*, 103–140. Retrieved from http://projekt.ht.lu.se/fileadmin/user_upload/project/grimm/working_papers/2009-jun/Hroarsdottir_WPSS83_notes.pdf

Paul, H. (1890). *Principles of the History of Language*. Рипол Классик.

Peters, P. (2009). *Comparative studies in Australian and New Zealand English: grammar and beyond*. Macquarie University.

Pinker, S. (1984). Visual cognition: An introduction. *Cognition*, *18*(1–3), 1–63. https://doi.org/10.1016/0010-0277(84)90021-0

Pinker, S. (1997). How the mind works. *Annals of the New York Academy of Sciences*, *882*(1), 119–

Pinker, S. (1998). Words and rules. Lingua 106 (1):219–42.

Pinker, S. (1999). Words and rules. The ingredients in language. London: Weidenfeld & Nicolson.

Pinker, S. (2001). Talk of genetics and vice versa. *Nature*, *413*(6855), 465–466. https://doi.org/10.1038/35097173

Pinker, S., & Mehler, J. (1988). *Connections and symbols* (Vol. 28). Mit Press.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, *28*(1), 73–193.

https://doi.org/10.1016/0010-0277(88)90032-7

Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, *6*(11), 456–463.

Plunkett, K., & Juola, P. (1999). A connectionist model of english past tense and plural morphology. *Cognitive Science*, *23*(4), 463–490. https://doi.org/10.1016/S0364-0213(99)00012-9

Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perception: Implications for child language acquisition. *Cognition*, *38*(1), 43–102. https://doi.org/10.1016/0010-0277(91)90022-V

Plunkett, K., & Marchman, V. (1993). From rote learning to system building: acquiring verb morphology in children and connectionist nets. *Cognition*, *48*(1), 21–69. https://doi.org/10.1016/0010-0277(93)90057-3

Postma, G. (2017). Modeling Transient States in Language Change. In *Micro-change and Macro-change in Diachronic Syntax*. Oxford Scholarship Online. https://doi.org/10.1093/oso/9780198747840.001.0001

Prasada, S., & Pinker, S. (1993). Generalisation of regular and irregular morphological patterns. *Language and Cognitive Processes*, *8*(1), 1–56. https://doi.org/10.1080/01690969308406948

Prasada, S., Pinker, S., & Snyder, W. (1990). Some evidence that irregular forms are retrieved from memory but regular forms are rule generated. Retrieved from https://philpapers.org/rec/PRASET

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. *Explorations in the Microstructure of Cognition*, *2*.

Say, T. (2000). *The mental representation of Italian morphology: evidence for the dual-mechanism model*. University of Essex.

Seidenberg, M. S., & Gonnerman, L. M. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences*, *4*(9), 353–361. https://doi.org/10.1016/S1364-6613(00)01515-1

Smolensky, P. (1996). On the Comprehension/Production Dilemma in Child Language. *Linguistic Inquiry*, *27*(4), 720–731. https://doi.org/10.2307/4178959

Sproat, R. W. (1992). *Morphology and computation*. MIT press.

Steele Jr, G. L. (1978). Rabbit: A compiler for Scheme.

Thomason, S. G. (2003). Contact as a Source of Language Change. In *The Handbook of Historical Linguistics* (pp. 687–712). Oxford, UK: Blackwell Publishing Ltd.

https://doi.org/10.1002/9780470756393.ch23

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, *74*(3), 209–253. https://doi.org/10.1016/S0010-0277(99)00069-4

Tomasello, M. (2003). Constructing a language: A usage-based account of language acquisition. *Cambridge, MA*.

Tomasello, M. (2009). The usage-based theory of language acquisition. In E. L. Bavin (Ed.), *The Cambridge Handbook of Child Language* (pp. 69–87). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511576164.005

Trudgill, P. (1986). *Dialects in contact*. Blackwell.

Weerman, F. (1989). *The V2 conspiracy : a synchronic and a diachronic analysis of verbal positions in Germanic languages*. Foris Publications. Retrieved from http://dare.uva.nl/search?arno.record.id=122592

Weerman, F. (1993). The diachronic consequences of first and second language acquisition: the change from OV to VO. *Linguistics*, *31*(5), 903–932. https://doi.org/10.1515/ling.1993.31.5.903

Weinreich, U., Labov, W., & Herzog, M. I. (1968). *Empirical foundations for a theory of language change*. University of Texas Press.

Willis, D. W. E. (1998). *Syntactic Change in Welsh: A Study of the Loss of the Verb-second*. Oxford University Press.

Xu, F., & Pinker, S. (1995). Weird past tense forms. *Journal of Child Language*, *22*(3), 531–556. https://doi.org/10.1017/S0305000900009946

Yang, C. (2002). *Knowledge and learning in natural language*. Oxford University Press on Demand.

Yang, C. (2005). On productivity. *Linguistic Variation Yearbook*, *5*(1), 265–302.

Yang, C. (2016). For and against frequencies. *Journal of Child Language*, *42*(2), 287–293. https://doi.org/10.1017/S0305000914000683

Zobl, H. (1998). Independent Representations of Verbal Affixes. *Morphology and Its Interfaces in Second Language Knowledge*, *19*, 339.

Appendices

Appendix 1: The 250 RVs and 250 IVs from WebCorp Corpus

	RVs			IVs	
	High frequency	Low frequency		High frequency	Low frequency
1	want	mouse	1	be	forgive
2	start	starve	2	have	overpay
3	use	oblige	3	do	thrust
4	call	redeploy	4	say	uphold
5	play	bewilder	5	get	creep
6	look	minimise	6	make	shine
7	ask	disgust	7	go	speed
8	love	criminalize	8	think	rewrite
9	post	trowel	9	come	mistake
10	decide	overstay	10	take	overtake
11	like	grub	11	tell	forecast
12	happen	brominate	12	see	mislead
13	try	frill	13	find	string
14	need	uplink	14	write	fling
15	seem	dizzy	15	give	undertake
16	work	federate	16	know	cling
17	turn	unsave	17	leave	shrink
18	add	revoke	18	put	weave
19	enjoy	blackmail	19	feel	withhold
20	miss	stratify	20	win	overthrow
21	finish	standardise	21	lose	stride
22	end	crack	22	hit	remake
23	receive	vouchsafe	23	become	plead
24	move	downconvert	24	buy	sweat
25	announce	code	25	begin	outgrow
26	help	crosspost	26	spend	kneel
27	mention	kettle	27	hear	outdo
28	sign	jacket	28	run	withstand
29	pick	unbundle	29	bring	redo
30	live	disconcert	30	keep	inset
31	stop	clog	31	set	breed
32	walk	defray	32	send	undo
33	include	furnish	33	fall	spell
34	watch	parcel	34	meet	forbid
35	realize	bush	35	read	podcast
36	die	snick	36	mean	beget
37	talk	privatise	37	throw	bend
38	drop	telescope	38	lead	offset

39	appear	cant	39	break	slit
40	report	frag	40	pay	recast
41	manage	train	41	catch	bust
42	allow	homogenise	42	choose	babysit
43	score	pearl	43	grow	befall
44	pass	band	44	speak	strive
45	create	upshift	45	beat	ride
46	open	retrograde	46	let	repay
47	notice	pink	47	hold	sling
48	fail	flitter	48	sit	foretell
49	pull	zombify	49	forget	tread
50	change	black	50	cut	outrun
51	follow	bullet	51	sell	dwell
52	continue	whelp	52	stand	wet
53	offer	employ	53	shoot	spoil
54	figure	contribute	54	drive	lean
55	arrive	prefabricate	55	wear	behold
56	vote	criticize	56	eat	override
57	release	certificate	57	draw	retell
58	return	revenge	58	teach	outshoot
59	explain	systematize	59	bet	spill
60	provide	dissolve	60	blow	overrun
61	comment	dignify	61	rise	bite
62	note	reconceive	62	strike	partake
63	agree	harrow	63	build	cleave
64	suggest	overpass	64	shut	retake
65	stay	branch	65	stick	oversleep
66	discover	quantify	66	wake	undercut
67	wonder	scarify	67	steal	underwrite
68	enter	abort	68	cost	ken
69	point	digitise	69	understand	waylay
70	join	theorise	70	quit	mishear
71	suffer	stimulate	71	hang	slink
72	reveal	perfect	72	fly	rewind
73	serve	victimize	73	fight	overspend
74	share	repress	74	sing	overshoot
75	dislike	overachieve	75	lie	wring
76	check	fictionalise	76	lay	inbreed
77	reach	immobilise	77	sleep	smite
78	save	depersonalize	78	cast	rethink
79	expect	unstate	79	seek	underthrow
80	laugh	shelter	80	wind	abide
81	visit	conjugate	81	dig	outshine
82	reply	recode	82	light	knit
83	raise	introduce	83	shake	unsay

84	deserve	externalize	84	hurt	bespeak
85	close	formalise	85	split	unwind
86	jump	synthesise	86	spread	redraw
87	kill	hairspray	87	ride	unbind
88	cause	involve	88	drink	forsake
89	respond	ornament	89	tear	unstick
90	describe	naturalise	90	upset	burn
91	claim	rewash	91	ring	cowrite
92	rank	suffix	92	shed	shoe
93	attend	winterize	93	shit	miscast
94	remind	snug	94	deal	forgo
95	feature	shrill	95	sweep	recut
96	state	computerise	96	bear	unmake
97	host	fog	97	hide	smell
98	launch	recycle	98	swing	foreknow
99	confirm	overdose	99	slide	typeset
100	file	categorize	100	freeze	inlay
101	sound	collocate	101	swear	intercut
102	believe	company	102	overcome	betake
103	complete	underperform	103	arise	dare
104	refuse	trackback	104	feed	heave
105	prove	suicide	105	learn	typecast
106	purchase	miscommunicate	106	undergo	rerun
107	name	eventuate	107	withdraw	overfeed
108	promise	trample	108	rebuild	uppercut
109	admit	recommence	109	spin	thrive
110	involve	witch	110	dive	overwrite
111	consider	crumb	111	sink	wed
112	order	leach	112	flee	sting
113	step	badge	113	sneak	chide
114	produce	roughen	114	lend	overblow
115	occur	stable	115	spit	gird
116	kick	slip	116	awake	overdraw
117	struggle	commercialise	117	bid	handwrite
118	average	winch	118	dream	unfreeze
119	wait	bogart	119	broadcast	bestrode
120	pitch	counterpunch	120	burst	bless
121	trade	globalise	121	swim	bereave
122	paint	unkink	122	stink	overhang
123	inspire	frank	123	weep	stave
124	head	unfit	124	spring	strip
125	carry	foray	125	bleed	clap
	250		Total	250	
Total	500				

Appendix 2: Word frequencies of IVs split by form and frequency from WebCorp Corpus

	Word frequencies of IVs with high frequency			Word frequencies of IVs with low frequency		
	Verbs	past	perfect	Verbs	past	perfect
1	be	4215057	1081787	forgive	506	2649
2	have	1147344	103941	overpay	497	1465
3	do	909206	243227	thrust	473	893
4	say	777450	51786	uphold	467	530
5	get	364219	105863	creep	455	472
6	make	248679	188063	shine	452	260
7	go	206213	85682	speed	425	246
8	think	196651	32005	rewrite	403	770
9	come	185412	77032	mistake	383	3914
10	take	170648	96814	overtake	377	713
11	tell	128587	38612	forecast	323	570
12	see	122744	170973	mislead	316	508
13	find	122430	68602	string	308	672
14	write	121799	74318	fling	307	387
15	give	98449	115429	undertake	305	788
16	know	97106	75836	cling	298	122
17	leave	87478	82422	shrink	271	810
18	put	79044	52353	weave	262	1453
19	feel	75681	12925	withhold	261	644
20	win	69927	30520	overthrow	258	256
21	lose	66879	63377	stride	253	0
22	hit	56947	28127	remake	243	533
23	become	52776	44622	plead	236	123
24	buy	51794	16025	sweat	232	0
25	begin	47992	10100	outgrow	228	517
26	spend	46682	30051	kneel	224	44
27	hear	43279	71973	outdo	224	934
28	run	40501	31119	withstand	198	80
29	bring	40364	25991	redo	185	515
30	keep	39339	15866	inset	173	23
31	set	37560	64778	breed	172	735
32	send	37424	22082	undo	171	1269
33	fall	31214	12189	spell	159	182
34	meet	29457	19988	forbid	159	1979
35	read	28890	63888	podcast	159	12
36	mean	28086	21320	beget	154	149
37	throw	27813	22642	bend	146	30
38	lead	27427	18560	offset	144	985
39	break	27145	16493	slit	141	51

40	pay	26451	57353	recast	139	261
41	catch	24811	26625	bust	136	1
42	choose	23605	16722	babysit	128	39
43	grow	24566	19345	befall	128	102
44	speak	23430	10132	strive	118	33
45	beat	23122	9086	ride	116	479
46	let	22822	5853	repay	107	1010
47	hold	22183	30376	sling	105	175
48	sit	21916	2972	foretell	98	102
49	forget	20897	15712	tread	97	136
50	cut	19169	16836	outrun	96	44
51	sell	19015	30959	dwell	89	47
52	stand	16503	1343	wet	87	0
53	shoot	13999	8638	spoil	85	375
54	drive	13967	10686	lean	84	29
55	wear	12504	8151	behold	84	28
56	eat	11673	6604	override	82	112
57	draw	11420	17099	retell	80	149
58	teach	11039	9893	outshoot	80	59
59	bet	10931	307	spill	79	172
60	blow	10226	11122	overrun	79	758
61	rise	10075	2788	bite	78	25
62	strike	8618	5538	partake	77	78
63	build	8004	29034	cleave	76	1
64	shut	7815	6568	retake	76	25
65	stick	7681	22001	oversleep	72	20
66	wake	7582	927	undercut	71	111
67	steal	7266	7592	underwrite	66	168
68	cost	7259	2269	ken	61	111
69	understand	7142	6253	waylay	58	106
70	quit	6725	1397	mishear	57	109
71	hang	6213	3116	slink	57	25
72	fly	6046	2074	rewind	56	11
73	fight	5747	3409	overspend	56	196
74	sing	5460	2252	overshoot	56	36
75	lie	5367	180	wring	50	90
76	lay	5174	10303	inbreed	48	83
77	sleep	4896	4	smite	46	1019
78	cast	4757	6060	rethink	44	68
79	seek	3878	3003	underthrow	35	4
80	wind	3781	461	abide	33	21
81	dig	3715	17	outshine	33	42
82	light	3665	3087	knit	31	3180
83	shake	3492	1646	unsay	31	137
84	hurt	3443	21903	bespeak	26	1

85	split	3405	3855	unwind	23	84
86	spread	3380	4026	redraw	23	0
87	ride	3184	863	unbind	22	30
88	drink	3103	1	forsake	22	380
89	tear	2774	5335	unstick	21	78
90	upset	2722	12730	burn	20	1204
91	ring	2279	1013	cowrite	19	6
92	shed	2199	2345	shoe	18	45
93	shit	2150	1618	miscast	14	166
94	deal	2110	5885	forgo	14	143
95	sweep	2081	3103	recut	13	10
96	bear	2053	31089	unmake	12	31
97	hide	1717	6245	smell	12	260
98	swing	1675	674	foreknow	11	5
99	slide	1569	175	typeset	11	39
100	freeze	1562	3139	inlay	11	30
101	swear	1533	2196	intercut	11	47
102	overcome	1514	1756	betake	11	3
103	arise	1507	384	dare	10	0
104	feed	1470	5994	heave	10	9
105	learn	1306	1415	typecast	9	131
106	undergo	1298	815	rerun	9	0
107	withdraw	1296	1651	overfeed	9	31
108	rebuild	1269	8	uppercut	9	3
109	spin	1240	1258	thrive	7	0
110	dive	1225	0	overwrite	7	121
111	sink	1212	1258	wed	7	653
112	flee	1104	337	sting	7	748
113	sneak	1098	383	chide	6	6
114	lend	1043	686	overblow	5	4
115	spit	1026	321	gird	5	7
116	awake	897	217	overdraw	4	240
117	bid	861	166	handwrite	4	0
118	dream	840	35	unfreeze	4	46
119	broadcast	759	2519	bestrode	4	0
120	burst	740	0	bless	3	52
121	swim	677	52	bereave	3	0
122	stink	671	259	overhang	3	7
123	weep	634	0	stave	2	1
124	spring	603	1130	strip	2	0
125	bleed	591	6	clap	2	0
		10,778,120	3,950,019		14,758	40,681
		14,728,139			55,439	
	Total	14,783,578				

Appendix 3: Word frequencies of RVs split by form and frequency from WebCorp Corpus

	Word frequencies of RVs with high frequency			Word frequencies of RVs with low frequency		
	Verbs	past	perfect	Verbs	past	perfect
1	want	147106	8097	mouse	5	3
2	start	125084	32829	starve	5	742
3	use	100311	126113	oblige	5	1639
4	call	82341	74400	redeploy	5	36
5	play	76218	51375	bewilder	5	680
6	look	75739	16112	minimise	5	12
7	ask	71364	37792	disgust	5	1711
8	love	68925	21331	criminalize	5	34
9	post	67393	28586	trowel	5	6
10	decide	64764	23367	overstay	5	54
11	like	61336	7907	grub	5	3
12	happen	58619	26758	brominate	5	2
13	try	56483	22469	frill	5	23
14	need	55218	32967	uplink	5	3
15	seem	54552	3299	dizzy	4	5
16	work	51125	29366	federate	4	42
17	turn	50952	25504	unsave	4	8
18	add	45504	18820	revoke	4	757
19	enjoy	45340	15323	blackmail	4	154
20	miss	43356	29708	stratify	4	68
21	finish	37342	16907	standardise	4	50
22	end	36110	11712	crack	4	2578
23	receive	35194	19136	vouchsafe	4	11
24	move	32903	19463	downconvert	4	0
25	announce	32625	24456	code	4	633
26	help	32052	14991	crosspost	4	14
27	mention	30136	29833	kettle	4	12
28	sign	29466	24788	jacket	4	2
29	pick	28465	15488	unbundle	4	7
30	live	26301	12204	disconcert	4	48
31	stop	26250	11711	clog	4	613
32	walk	26173	5	defray	4	3
33	include	26157	16392	furnish	4	534
34	watch	25960	13046	parcel	4	8
35	realize	24844	5427	bush	4	25
36	die	24260	6358	snick	4	0
37	talk	23185	11201	privatise	4	20
38	drop	23134	10010	telescope	4	9
39	appear	23111	3758	cant	4	15

40	report	22088	13220	frag	4	8
41	manage	21760	14351	train	4	5677
42	allow	21376	41752	homogenise	4	17
43	score	21325	6712	pearl	4	33
44	pass	21251	20716	band	4	291
45	create	20602	34103	upshift	4	0
46	open	20479	7414	retrograde	4	0
47	notice	20030	14670	pink	4	7
48	fail	19525	10408	flitter	4	12
49	pull	19152	9914	zombify	4	26
50	change	18190	33578	black	4	890
51	follow	17095	16910	bullet	4	5
52	continue	17066	5833	whelp	4	1
53	offer	17063	18712	employ	4	7657
54	figure	16802	4816	contribute	4	2930
55	arrive	16595	6925	prefabricate	4	61
56	vote	16545	5756	criticize	4	3290
57	release	16538	47306	certificate	4	14
58	return	16418	9396	revenge	4	9
59	explain	16271	4367	systematize	4	29
60	provide	16197	19719	dissolve	4	930
61	comment	15281	26	dignify	4	0
62	note	15154	6961	reconceive	4	20
63	agree	15120	7376	harrow	4	8
64	suggest	14785	5781	overpass	4	0
65	stay	14246	4817	branch	4	283
66	discover	14109	10505	quantify	3	165
67	wonder	14026	10	scarify	3	6
68	enter	13932	8207	abort	3	251
69	point	13863	4896	digitise	3	17
70	join	13567	8267	theorise	3	6
71	suffer	12917	5398	stimulate	3	524
72	reveal	12398	10807	perfect	3	1386
73	serve	12286	11930	victimize	3	705
74	share	12241	12361	repress	3	184
75	dislike	12189	505	overachieve	3	68
76	check	12139	6468	fictionalise	3	23
77	reach	12112	13370	immobilise	3	12
78	save	11955	17213	depersonalize	3	7
79	expect	11542	47541	unstate	3	24
80	laugh	11479	2180	shelter	3	321
81	visit	11308	4993	conjugate	3	9
82	reply	11262	418	recode	3	2
83	raise	11245	19483	introduce	3	9719
84	deserve	11164	113	externalize	3	24

85	close	10850	19092	formalise	3	21
86	jump	10796	1542	synthesise	3	6
87	kill	10670	15507	hairspray	3	0
88	cause	10664	10360	involve	3	46031
89	respond	10503	2399	ornament	3	20
90	describe	10461	9368	naturalise	3	9
91	claim	10314	3962	rewash	3	1
92	rank	10273	13821	suffix	3	4
93	attend	10203	2637	winterize	3	49
94	remind	10156	7593	snug	3	11
95	feature	10083	10805	shrill	3	0
96	state	10061	9253	computerise	3	41
97	host	9996	177	fog	3	0
98	launch	9973	8300	recycle	3	932
99	confirm	9916	9939	overdose	3	119
100	file	9814	14434	categorize	3	720
101	sound	9595	885	collocate	3	8
102	believe	9386	6995	company	3	0
103	complete	9347	12186	underperform	3	348
104	refuse	9291	2328	trackback	3	3
105	prove	9087	4293	suicide	3	6
106	purchase	8951	10224	miscommunicate	3	6
107	name	8923	33108	eventuate	3	8
108	promise	8232	5465	trample	3	510
109	admit	8231	4518	recommence	3	0
110	involve	8217	46031	witch	3	2
111	consider	8195	36835	crumb	3	4
112	order	8149	4959	leach	3	25
113	step	8145	2942	badge	3	8
114	produce	8094	12492	roughen	3	6
115	occur	8052	3505	stable	3	9
116	kick	7964	6006	slip	3	1462
117	struggle	7907	5750	commercialise	3	22
118	average	7725	1368	winch	3	9
119	wait	7712	2915	bogart	2	4
120	pitch	7678	5079	counterpunch	2	1
121	trade	7638	10618	globalise	2	6
122	paint	7574	8083	unkink	2	0
123	inspire	7569	20294	frank	2	9
124	head	7519	14907	unfit	2	4
125	carry	7513	6092	foray	2	5
Total		3,019,513	1,862,280		447	96,644
		4,881,793			97,091	
		4,978,884				

Appendix 4: Word frequencies of RFs split by form and frequency from WebCorp Corpus

	Word frequencies of IVs with high frequency			Word frequencies of IVs with low frequency		
	Verbs	past	perfect	Verbs	past	perfect
1	be	0	0	forgive	2	2
2	have	19	3	overpay	9	14
3	do	0	0	thrust	10	8
4	say	7	3	uphold	0	0
5	get	0	0	creep	233	427
6	make	14	2	shine	446	269
7	go	3	0	speed	75	5
8	think	10	2	rewrite	0	0
9	come	1	2	mistake	2	1
10	take	7	7	overtake	0	1
11	tell	0	1	forecast	55	205
12	see	0	0	mislead	0	1
13	find	3	1	string	0	0
14	write	4	8	fling	0	0
15	give	6	4	undertake	0	0
16	know	8	2	cling	4	0
17	leave	0	0	shrink	2	0
18	put	30	22	weave	193	126
19	feel	6	1	withhold	0	0
20	win	2	1	overthrow	0	0
21	lose	5	1	stride	2	1
22	hit	3	0	remake	0	2
23	become	1	2	plead	1510	382
24	buy	7	3	sweat	12	249
25	begin	0	0	outgrow	0	0
26	spend	4	0	kneel	41	18
27	hear	38	23	outdo	0	0
28	run	1	2	withstand	0	0
29	bring	0	1	redo	0	0
30	keep	1	1	inset	0	0
31	set	1	2	breed	0	0
32	send	3	1	stink	20	2
33	fall	5	6	undo	0	0
34	meet	0	0	spell	92	3094
35	read	9	2	forbid	0	0
36	mean	3	0	podcast	7	8
37	throw	3	5	bend	15	5
38	lead	0	0	offset	0	0
39	break	3	2	slit	0	0

40	pay	631	526	recast	7	15
41	catch	20	11	bust	1308	1327
42	choose	22	2	babysit	2	0
43	grow	15	20	befall	0	1
44	speak	0	0	strive	62	107
45	beat	8	5	rid	4	3
46	let	1	0	repay	4	7
47	hold	5	0	sling	0	0
48	sit	1	1	foretell	0	0
49	forget	0	2	tread	42	9
50	cut	3	0	outrun	0	0
51	sell	0	0	dwell	46	25
52	stand	3	1	wet	6	20
53	shoot	6	0	spoil	1311	0
54	drive	6	4	lean	31643	447
55	wear	1	0	behold	0	0
56	eat	5	4	override	0	0
57	draw	9	3	retell	0	0
58	teach	11	5	outshoot	0	0
59	bet	5	3	spill	1147	707
60	blow	0	58	overrun	0	0
61	rise	4	3	bite	1	0
62	strike	6	7	partake	0	0
63	build	6	8	cleave	7	27
64	shut	0	0	retake	0	0
65	stick	20	10	oversleep	0	0
66	wake	9	6	undercut	0	0
67	steal	0	0	underwrite	0	0
68	cost	55	84	ken	0	2
69	understand	0	1	waylay	0	1
70	quit	3	3	mishear	0	0
71	hang	15	405	slink	46	10
72	fly	0	0	rewind	5	0
73	fight	2	0	overspend	0	0
74	sing	0	0	overshoot	0	0
75	lie	0	0	wring	0	0
76	lay	167	180	inbreed	0	0
77	sleep	2	4	smite	9	12
78	cast	34	21	rethink	0	0
79	seek	6	8	underthrow	0	0
80	wind	0	0	abide	22	19
81	dig	13	8	outshine	40	28
82	light	4	209	knit	33	1189
83	shake	4	4	unsay	0	0
84	hurt	9	1	bespeak	0	0

85	split	1	1	unwind	0	0
86	spread	2	6	redraw	0	0
87	ride	1	0	unbind	0	0
88	drink	1	0	forsake	2	0
89	tear	0	0	unstick	0	0
90	upset	0	0	burn	2328	5399
91	ring	0	0	cowrite	0	0
92	shed	8	2	shoe	8	14
93	shit	2	5	miscast	1	0
94	deal	3	18	forgo	2	0
95	sweep	1	1	recut	0	0
96	bear	5	5	unmake	0	0
97	hide	3	0	smell	1742	4
98	swing	1	3	foreknow	0	0
99	slide	6	3	typeset	0	0
100	freeze	2	6	inlay	1	3
101	swear	1	1	intercut	0	0
102	overcome	0	1	betake	0	0
103	arise	2	3	dare	1518	9
104	feed	0	1	heave	198	49
105	learn	31643	24408	typecast	0	0
106	undergo	0	0	rerun	0	0
107	withdraw	0	0	overfeed	0	0
108	rebuild	0	0	uppercut	1	0
109	spin	0	0	thrive	761	0
110	dive	339	0	overwrite	0	1
111	sink	3	1	wed	0	313
112	flee	0	0	sting	0	0
113	sneak	485	1	chide	174	85
114	lend	6	4	overblow	0	0
115	spit	2	2	gird	39	2
116	awake	3	1	overdraw	0	0
117	bid	0	0	handwrite	0	0
118	dream	1923	1513	unfreeze	0	0
119	broadcast	68	134	bestrode	0	0
120	burst	12	7	bless	56	17417
121	swim	0	0	bereave	10	103
122	weep	7	0	overhang	1	0
123	spring	0	1	stave	77	18
124	bleed	0	0	strip	1202	2494
125	leap	500	67	clap	245	142
Total	36348		27912	46841		34829
	64260			81670		
	145930					

Appendix 5: Word frequencies of IVs (without suppletives) split by form and frequency from WebCorp Corpus

	Word frequencies of IVs with high frequency			Word frequencies of IVs with low frequency		
	verbs	past	perfect	verbs	past	perfect
1	say	777450	51786	overpay	497	1465
2	get	364219	105863	thrust	473	893
3	make	248679	188063	uphold	467	530
4	think	196651	32005	creep	455	472
5	come	185412	77032	shine	452	260
6	take	170648	96814	speed	425	246
7	tell	128587	38612	rewrite	403	770
8	see	122744	170973	mistake	383	3914
9	find	122430	68602	overtake	377	713
10	write	121799	74318	pen	336	0
11	give	98449	115429	forecast	323	570
12	know	97106	75836	mislead	316	508
13	leave	87478	82422	string	308	672
14	put	79044	52353	fling	307	387
15	feel	75681	12925	undertake	305	788
16	win	69927	30520	cling	298	122
17	lose	66879	63377	shrink	271	810
18	hit	56947	28127	weave	262	1453
19	become	52776	44622	withhold	261	644
20	buy	51794	16025	overthrow	258	256
21	begin	47992	10100	stride	253	0
22	spend	46682	30051	remake	243	533
23	hear	43279	71973	plead	236	123
24	run	40501	31119	sweat	232	0
25	bring	40364	25991	outgrow	228	517
26	keep	39339	15866	kneel	224	44
27	set	37560	64778	withstand	198	80
28	send	37424	22082	inset	173	23
29	fall	31214	12189	foresee	172	240
30	meet	29457	19988	breed	172	735
31	read	28890	63888	stink	170	259
32	mean	28086	21320	spell	159	182
33	throw	27813	22642	forbid	159	1979
34	lead	27427	18560	podcast	159	12
35	break	27145	16493	bend	146	30
36	pay	26451	57353	offset	144	985
37	catch	24811	26625	slit	141	51

38	choose	23605	16722	recast	139	261
39	grow	24566	19345	bust	136	1
40	speak	23430	10132	babysit	128	39
41	beat	23122	9086	befall	128	102
42	let	22822	5853	strive	118	33
43	hold	22183	30376	rid	116	479
44	sit	21916	2972	repay	107	1010
45	forget	20897	15712	sling	105	175
46	cut	19169	16836	foretell	98	102
47	sell	19015	30959	tread	97	136
48	stand	16503	1343	outrun	96	44
49	shoot	13999	8638	dwell	89	47
50	drive	13967	10686	wet	87	0
51	wear	12504	8151	spoil	85	375
52	eat	11673	6604	lean	84	29
53	draw	11420	17099	behold	84	28
54	teach	11039	9893	override	82	112
55	bet	10931	307	retell	80	149
56	blow	10226	11122	outshoot	80	59
57	rise	10075	2788	spill	79	172
58	strike	8618	5538	overrun	79	758
59	build	8004	29034	bite	78	1265
60	shut	7815	6568	partake	77	78
61	stick	7681	22001	cleave	76	1
62	wake	7582	927	retake	76	25
63	stole	7266	7592	oversleep	72	20
64	cost	7259	2269	undercut	71	111
65	quit	6725	1397	underwrite	66	168
66	understand	7142	6253	ken	61	111
67	hang	6213	3116	waylay	58	106
68	fly	6046	2074	mishear	57	109
69	fight	5747	3409	slink	57	25
70	sing	5460	2252	rewind	56	11
71	lie	5367	180	overspend	56	196
72	lay	5174	10303	overshoot	56	36
73	sleep	4896	4	wring	50	90
74	cast	4757	6060	inbreed	48	83
75	seek	3878	3003	smite	46	1019
76	wind	3781	461	rethink	44	68
77	dig	3715	17	beget	41	149
78	light	3665	3087	underthrow	35	4
79	shake	3492	1646	abide	33	21
80	hurt	3443	21903	outshine	33	42
81	split	3405	3855	knit	31	3180
82	spread	3380	4026	unsay	31	137

83	ride	3184	863	bespeak	26	1
84	drink	3103	1	unwind	23	84
85	tear	2774	5335	redraw	23	0
86	upset	2722	12730	unbind	22	30
87	ring	2279	1013	forsake	22	380
88	shed	2199	2345	unstick	21	78
89	shit	2150	1618	burnt	20	1204
90	deal	2110	5885	cowrite	19	6
91	sweep	2081	3103	shoe	18	45
92	bear	2053	31089	miscast	14	166
93	hide	1717	6245	recut	13	10
94	swing	1675	674	unmake	12	31
95	slide	1569	175	smell	12	260
96	freeze	1562	3139	foreknow	11	5
97	swear	1533	2196	typeset	11	39
98	overcome	1514	1756	inlay	11	30
99	arise	1507	384	intercut	11	47
100	feed	1470	5994	betake	11	3
101	learn	1306	1415	heave	10	9
102	withdraw	1296	1651	typecast	9	131
103	rebuild	1269	8	rerun	9	0
104	spin	1240	1258	overfeed	9	31
105	sink	1212	1258	uppercut	9	3
106	flee	1104	337	thrive	7	0
107	sneak	1098	383	overwrite	7	121
108	lend	1043	686	wed	7	653
109	spit	1026	321	sting	7	746
110	awake	897	217	chide	6	6
111	bid	861	166	overblow	5	4
112	dream	840	35	gird	5	7
113	broadcast	759	2519	overdraw	4	240
114	burst	740	0	handwrite	4	0
115	oversee	699	538	unfreeze	4	46
116	swim	677	52	bestride	4	0
117	weep	634	0	bless	3	52
118	spring	603	1130	bereave	3	0
119	bleed	591	6	overhang	3	7
120	grind	579	589	stave	2	1
121	leap	544	3	strip	2	0
122	forgive	506	2649	clap	2	0
	Total	4299434	2438087		14213	36908
		6737521			51121	
		6788642				

Appendix 6: Word frequencies of IVs and IFs with high frequency in the past and perfect forms from WebCorp Corpus

Verbs	Word freq.	Form	Frequency	Class	Class -N.	Vowel -N.	Class-change	Class-N.	Vowel-N.	IFs	Word freq.	Direction of vowel change
say	777450	past	High	2C-6	24	2	0	0	0	0	0	0
get	364219	past	High	2D-5	32	2	0	0	0	0	0	0
make	248679	past	High	1C-5	7	1	0	0	0	0	0	0
think	196651	past	High	2C-9	27	2	0	0	0	0	0	0
come	185412	past	High	2A-1	10	2	0	0	0	0	0	0
take	170648	past	High	2B-1	12	2	0	0	0	0	0	0
tell	128587	past	High	2C-8	26	2	0	0	0	0	0	0
see	122744	past	High	2B-5	16	2	0	0	0	0	0	0
find	122430	past	High	2C-5	23	2	0	0	0	0	0	0
write	121799	past	High	3A-1	33	3	0	0	0	0	0	0
give	98449	past	High	2B-2	13	2	0	0	0	0	0	0
know	97106	past	High	2B-3	14	2	0	0	0	0	0	0
leave	87478	past	High	2C-6	24	2	0	0	0	0	0	0
put	79044	past	High	1A-1	1	1	0	0	0	0	0	0
feel	75681	past	High	2C-6	24	2	0	0	0	0	0	0
win	69927	past	High	2C-1	19	2	0	0	0	0	0	0
lose	66879	past	High	2C-7	25	2	0	0	0	0	0	0
hit	56947	past	High	1A-1	1	1	0	0	0	0	0	0
become	52776	past	High	2A-1	10	2	0	0	0	0	0	0
buy	51794	past	High	2C-9	27	2	0	0	0	0	0	0
begin	47992	past	High	3A-2	34	3	0	0	0	0	0	0
spend	46682	past	High	1C-6	8	1	0	0	0	0	0	0
hear	43279	past	High	2C-4	22	2	0	0	0	0	0	0
run	40501	past	High	2A-2	11	2	0	0	0	0	0	0
bring	40364	past	High	2C-9	27	2	2C-1	19	2	brung	27	0
keep	39339	past	High	2C-6	24	2	0	0	0	0	0	0
set	37560	past	High	1A-1	1	1	0	0	0	0	0	0
send	37424	past	High	1C-6	8	1	0	0	0	0	0	0
fall	31214	past	High	2B-7	18	2	0	0	0	0	0	0
meet	29457	past	High	2C-6	24	2	0	0	0	0	0	0
read	28890	past	High	2C-6	24	2	0	0	0	0	0	0
mean	28086	past	High	2C-6	24	2	0	0	0	0	0	0
throw	27813	past	High	2B-3	14	2	0	0	0	0	0	0
lead	27427	past	High	2C-6	24	2	0	0	0	0	0	0
break	27145	past	High	2D-4	31	2	0	0	0	0	0	0
pay	26451	past	High	1C-7	9	1	0	0	0	0	0	0
catch	24811	past	High	2C-9	27	2	0	0	0	0	0	0
choose	23605	past	High	2D-4	31	2	0	0	0	0	0	0
grow	24566	past	High	2B-3	14	2	0	0	0	0	0	0
speak	23430	past	High	2D-4	31	2	0	0	0	0	0	0
beat	23122	past	High	1B-1	2	1	0	0	0	0	0	0
let	22822	past	High	1A-1	1	1	0	0	0	0	0	0
hold	22183	past	High	2C-6	24	2	0	0	0	0	0	0
sit	21916	past	High	2C-2	20	2	0	0	0	0	0	0
forget	20897	past	High	2D-5	32	2	0	0	0	0	0	0

cut	19169	past	High	1A-1	1	1	0	0	0	0	0	0
sell	19015	past	High	2C-8	26	2	0	0	0	0	0	0
stand	16503	past	High	2C-3	21	2	0	0	0	0	0	0
shoot	13999	past	High	2C-7	25	2	0	0	0	0	0	0
drive	13967	past	High	3A-1	33	3	0	0	0	0	0	0
wear	12504	past	High	2D-1	28	2	0	0	0	0	0	0
eat	11673	past	High	2B-6	17	2	0	0	0	0	0	0
draw	11420	past	High	2B-4	15	2	0	0	0	0	0	0
teach	11039	past	High	2C-9	27	2	0	0	0	0	0	0
bet	10931	past	High	1A-1	1	1	0	0	0	0	0	0
blow	10226	past	High	2B-3	14	2	0	0	0	0	0	0
rise	10075	past	High	3A-1	33	3	0	0	0	0	0	0
strike	8618	past	High	2C-1	19	2	0	0	0	0	0	0
build	8004	past	High	1C-6	8	1	0	0	0	0	0	0
shut	7815	past	High	1A-1	1	1	0	0	0	0	0	0
stick	7681	past	High	2C-1	19	2	0	0	0	0	0	0
wake	7582	past	High	2D-4	31	2	0	0	0	0	0	0
steal	7266	past	High	2D-4	31	2	0	0	0	0	0	0
cost	7259	past	High	1A-1	1	1	0	0	0	0	0	0
quit	6725	past	High	1A-1	1	1	0	0	0	0	0	0
understand	7142	past	High	2C-3	21	2	0	0	0	0	0	0
hang	6213	past	High	2C-1	19	2	0	0	0	0	0	0
fly	6046	past	High	3A-3	35	3	0	0	0	0	0	0
fight	5747	past	High	2C-9	27	2	0	0	0	0	0	0
sing	5460	past	High	3A-2	34	3	2C-1	19	2	sung	3	decrease
lie	5367	past	High	2D-3	30	2	0	0	0	0	0	0
lay	5174	past	High	1C-7	9	1	0	0	0	0	0	0
sleep	4896	past	High	2C-6	24	2	0	0	0	0	0	0
cast	4757	past	High	1A-1	1	1	0	0	0	0	0	0
seek	3878	past	High	2C-9	27	3	0	0	0	0	0	0
wind	3781	past	High	2C-5	23	2	0	0	0	0	0	0
dig	3715	past	High	2C-1	19	2	0	0	0	0	0	0
light	3665	past	High	2D-2	29	2	0	0	0	0	0	0
shake	3492	past	High	2B-1	12	2	0	0	0	0	0	0
hurt	3443	past	High	1A-1	1	1	0	0	0	0	0	0
split	3405	past	High	1A-1	1	1	0	0	0	0	0	0
spread	3380	past	High	1A-1	1	1	0	0	0	0	0	0
ride	3184	past	High	3A-1	33	3	0	0	0	0	0	0
drink	3103	past	High	3A-2	34	3	0	0	0	0	0	0
tear	2774	past	High	2D-1	28	2	0	0	0	0	0	0
upset	2722	past	High	1A-1	1	1	0	0	0	0	0	0
ring	2279	past	High	3A-2	34	3	2C-1	19	2	rung	2	decrease
shed	2199	past	High	1A-1	1	1	0	0	0	0	0	0
shit	2150	past	High	1A-1	1	1	2C-2	20	2	shat	239	increase
deal	2110	past	High	2C-6	24	2	0	0	0	0	0	0
sweep	2081	past	High	2C-6	24	2	0	0	0	0	0	0
bear	2053	past	High	2D-1	28	2	0	0	0	0	0	0
hide	1717	past	High	2D-2	29	2	0	0	0	0	0	0
swing	1675	past	High	2C-1	19	2	3A-2	34	3	swang	1	increase
slide	1569	past	High	2D-2	29	2	0	0	0	0	0	0
freeze	1562	past	High	2D-4	31	2	0	0	0	0	0	0
swear	1533	past	High	2D-1	28	2	0	0	0	0	0	0
overcome	1514	past	High	2A-1	10	2	0	0	0	0	0	0

arise	1507	past	High	3A-1	33	3	0	0	0	0	0	0
feed	1470	past	High	2C-6	24	2	0	0	0	0	0	0
learn	1306	past	High	1C-1	3	1	0	0	0	0	0	0
withdraw	1296	past	High	2B-4	15	2	0	0	0	0	0	0
rebuild	1269	past	High	1C-6	8	1	0	0	0	0	0	0
spin	1240	past	High	2C-1	19	2	0	0	0	0	0	0
sink	1212	past	High	3A-2	34	3	2C-1	19	2	sunk	588	decrease
flee	1104	past	High	2C-6	24	2	0	0	0	0	0	0
sneak	1098	past	High	2C-1	19	2	0	0	0	0	0	0
lend	1043	past	High	1C-6	8	1	0	0	0	0	0	0
spit	1026	past	High	1A-1	1	1	2C-2	20	2	spat	394	increase
awake	897	past	High	2D-4	31	2	0	0	0	0	0	0
bid	861	past	High	1A-1	1	1	0	0	0	0	0	0
dream	840	past	High	2C-6	24	2	0	0	0	0	0	0
broadcast	759	past	High	1A-1	1	1	0	0	0	0	0	0
burst	740	past	High	1A-1	1	1	0	0	0	0	0	0
oversee	699	past	High	2B-5	16	2	0	0	0	0	0	0
swim	677	past	High	3A-2	34	3	0	0	0	0	0	0
weep	634	past	High	2C-6	24	2	0	0	0	0	0	0
spring	603	past	High	3A-2	34	3	2C-1	19	2	sprung	3	decrease
bleed	591	past	High	2C-6	24	2	0	0	0	0	0	0
grind	579	past	High	2C-5	23	2	0	0	0	0	0	0
leap	544	past	High	2C-6	24	2	0	0	0	0	0	0
forgive	506	past	High	2B-2	13	2	0	0	0	0	0	0
say	51786	perfect	High	2C-6	24	2	0	0	0	0	0	0
get	105863	perfect	High	2D-5	32	1	0	0	0	0	0	0
make	188063	perfect	High	1C-5	7	1	0	0	0	0	0	0
think	32005	perfect	High	2C-9	27	2	0	0	0	0	0	0
come	77032	perfect	High	2A-1	10	2	0	0	0	0	0	0
take	96814	perfect	High	2B-1	12	2	0	0	0	0	0	0
tell	38612	perfect	High	2C-8	26	2	0	0	0	0	0	0
see	170973	perfect	High	2B-5	16	2	0	0	0	0	0	0
find	68602	perfect	High	2C-5	23	2	0	0	0	0	0	0
write	74318	perfect	High	3A-1	33	3	0	0	0	0	0	0
give	115429	perfect	High	2B-2	13	2	0	0	0	0	0	0
know	75836	perfect	High	2B-3	14	2	0	0	0	0	0	0
leave	82422	perfect	High	2C-6	24	2	0	0	0	0	0	0
put	52353	perfect	High	1A-1	1	1	0	0	0	0	0	0
feel	12925	perfect	High	2C-6	24	2	0	0	0	0	0	0
win	30520	perfect	High	2C-1	19	2	0	0	0	0	0	0
lose	63377	perfect	High	2C-7	25	2	0	0	0	0	0	0
hit	28127	perfect	High	1A-1	1	1	0	0	0	0	0	0
become	44622	perfect	High	2A-1	10	2	0	0	0	0	0	0
buy	16025	perfect	High	2C-9	27	2	0	0	0	0	0	0
begin	10100	perfect	High	3A-2	34	3	0	0	0	0	0	0
spend	30051	perfect	High	1C-6	8	1	0	0	0	0	0	0
hear	71973	perfect	High	2C-4	22	2	0	0	0	0	0	0
run	31119	perfect	High	2A-2	11	2	0	0	0	0	0	0
bring	25991	perfect	High	2C-9	27	2	2C-1	19	2	brung	5	0
keep	15866	perfect	High	2C-6	24	2	0	0	0	0	0	0
set	64778	perfect	High	1A-1	1	1	0	0	0	0	0	0
send	22082	perfect	High	1C-6	8	1	0	0	0	0	0	0
fall	12189	perfect	High	2B-7	18	2	0	0	0	0	0	0

meet	19988	perfect	High	2C-6	24	2	0	0	0	0	0	0
read	63888	perfect	High	2C-6	24	2	0	0	0	0	0	0
mean	21320	perfect	High	2C-6	24	2	0	0	0	0	0	0
throw	22642	perfect	High	2B-3	14	2	0	0	0	0	0	0
lead	18560	perfect	High	2C-6	24	2	0	0	0	0	0	0
break	16493	perfect	High	2D-4	31	2	0	0	0	0	0	0
pay	57353	perfect	High	1C-7	9	1	0	0	0	0	0	0
catch	26625	perfect	High	2C-9	27	2	0	0	0	0	0	0
choose	16722	perfect	High	2D-4	31	2	0	0	0	0	0	0
grow	19345	perfect	High	2B-3	14	2	0	0	0	0	0	0
speak	10132	perfect	High	2D-4	31	2	0	0	0	0	0	0
beat	9086	perfect	High	1B-1	2	1	0	0	0	0	0	0
let	5853	perfect	High	1A-1	1	1	0	0	0	0	0	0
hold	30376	perfect	High	2C-6	24	2	0	0	0	0	0	0
sit	2972	perfect	High	2C-2	20	2	0	0	0	0	0	0
forget	15712	perfect	High	2D-5	32	2	2C-7	25	2	forgot	2798	0
cut	16836	perfect	High	1A-1	1	1	0	0	0	0	0	0
sell	30959	perfect	High	2C-8	26	2	0	0	0	0	0	0
stand	1343	perfect	High	2C-3	21	2	0	0	0	0	0	0
shoot	8638	perfect	High	2C-7	25	2	0	0	0	0	0	0
drive	10686	perfect	High	3A-1	33	3	0	0	0	0	0	0
wear	8151	perfect	High	2D-1	28	2	0	0	0	0	0	0
eat	6604	perfect	High	2B-6	17	2	0	0	0	0	0	0
draw	17099	perfect	High	2B-4	15	2	0	0	0	0	0	0
teach	9893	perfect	High	2C-9	27	2	0	0	0	0	0	0
bet	307	perfect	High	1A-1	1	1	0	0	0	0	0	0
blow	11122	perfect	High	2B-3	14	2	0	0	0	0	0	0
rise	2788	perfect	High	3A-1	33	3	0	0	0	0	0	0
strike	5538	perfect	High	2C-1	19	2	3A-1	33	3	stricken	546	increase
build	29034	perfect	High	1C-6	8	1	0	0	0	0	0	0
shut	6568	perfect	High	1A-1	1	1	0	0	0	0	0	0
stick	22001	perfect	High	2C-1	19	2	0	0	0	0	0	0
wake	927	perfect	High	2D-4	31	2	0	0	0	0	0	0
steal	7592	perfect	High	2D-4	31	2	0	0	0	0	0	0
cost	2269	perfect	High	1A-1	1	1	0	0	0	0	0	0
quit	1397	perfect	High	1A-1	1	1	0	0	0	0	0	0
understand	6253	perfect	High	2C-3	21	2	0	0	0	0	0	0
hang	3116	perfect	High	2C-1	19	2	0	0	0	0	0	0
fly	2074	perfect	High	3A-3	35	3	0	0	0	0	0	0
fight	3409	perfect	High	2C-9	27	2	0	0	0	0	0	0
sing	2252	perfect	High	3A-2	34	3	0	0	0	0	0	0
lie	180	perfect	High	2D-3	30	2	0	0	0	0	0	0
lay	10303	perfect	High	1C-7	9	1	0	0	0	0	0	0
sleep	4	perfect	High	2C-6	24	2	0	0	0	0	0	0
cast	6060	perfect	High	1A-1	1	1	0	0	0	0	0	0
seek	3003	perfect	High	2C-9	27	2	0	0	0	0	0	0
wind	461	perfect	High	2C-5	23	2	0	0	0	0	0	0
dig	17	perfect	High	2C-1	19	2	0	0	0	0	0	0
light	3087	perfect	High	2D-2	29	2	0	0	0	0	0	0
shake	1646	perfect	High	2B-1	12	2	0	0	0	0	0	0
hurt	21903	perfect	High	1A-1	1	1	0	0	0	0	0	0
split	3855	perfect	High	1A-1	1	1	0	0	0	0	0	0
spread	4026	perfect	High	1A-1	1	1	0	0	0	0	0	0

ride	863	perfect	High	3A-1	33	3	0	0	0	0	0	0
drink	1	perfect	High	3A-2	34	3	0	0	0	0	0	0
tear	5335	perfect	High	2D-1	28	2	0	0	0	0	0	0
upset	12730	perfect	High	1A-1	1	1	0	0	0	0	0	0
ring	1013	perfect	High	3A-2	33	3	0	0	0	0	0	0
shed	2345	perfect	High	1A-1	1	1	0	0	0	0	0	0
shit	1618	perfect	High	1A-1	1	1	0	0	0	0	0	0
deal	5885	perfect	High	2C-6	24	2	0	0	0	0	0	0
sweep	3103	perfect	High	2C-6	24	2	0	0	0	0	0	0
bear	31089	perfect	High	2D-1	28	2	0	0	0	0	0	0
hide	6245	perfect	High	2D-2	29	2	0	0	0	0	0	0
swing	674	perfect	High	2C-1	19	2	0	0	0	0	0	0
slide	175	perfect	High	2D-2	29	2	0	0	0	0	0	0
freeze	3139	perfect	High	2D-4	31	2	0	0	0	0	0	0
swear	2196	perfect	High	2D-1	28	2	0	0	0	0	0	0
overcome	1756	perfect	High	2A-1	10	2	0	0	0	0	0	0
arise	384	perfect	High	3A-1	33	3	0	0	0	0	0	0
feed	5994	perfect	High	2C-6	24	2	0	0	0	0	0	0
learn	1415	perfect	High	1C-1	3	1	0	0	0	0	0	0
withdraw	1651	perfect	High	2B-4	15	2	0	0	0	0	0	0
rebuild	8	perfect	High	1C-6	8	1	0	0	0	0	0	0
spin	1258	perfect	High	2C-1	19	2	0	0	0	0	0	0
sink	1258	perfect	High	3A-2	34	3	0	0	0	0	0	0
flee	337	perfect	High	2C-6	24	2	0	0	0	0	0	0
sneak	383	perfect	High	2C-1	19	2	0	0	0	0	0	0
lend	686	perfect	High	1C-6	8	1	0	0	0	0	0	0
spit	321	perfect	High	1A-1	1	1	2C-2	20	2	spat	142	increase
awake	217	perfect	High	2D-4	31	2	0	0	0	0	0	0
bid	166	perfect	High	1A-1	1	1	0	0	0	0	0	0
dream	35	perfect	High	2C-6	24	2	0	0	0	0	0	0
broadcast	2519	perfect	High	1A-1	1	1	0	0	0	0	0	0
burst	0	perfect	High	1A-1	1	1	0	0	0	0	0	0
oversee	538	perfect	High	2B-5	16	2	0	0	0	0	0	0
swim	52	perfect	High	3A-2	34	3	0	0	0	0	0	0
weep	0	perfect	High	2C-6	24	2	0	0	0	0	0	0
spring	1130	perfect	High	3A-2	34	3	0	0	0	0	0	0
bleed	6	perfect	High	2C-6	24	2	0	0	0	0	0	0
grind	589	perfect	High	2C-5	23	2	0	0	0	0	0	0
leap	3	perfect	High	2C-6	24	2	0	0	0	0	0	0
forgive	2649	perfect	High	2B-2	13	2	0	0	0	0	0	0
Total	6,737,521										4,748	

Appendix 7: Word frequencies of IVs and IFs with low frequency in the past and perfect forms from WebCorp Corpus

Verbs	Word freq.	Form	Frequency	Class	Class-N.	Vowel-N.	Class-change	Class-N.	Vowel-N.	IFs	Word freq.	Direction of vowel change
overpay	497	past	Low	1C-7	9	1	0	0	0	0	0	0
thrust	473	past	Low	1A-1	1	1	0	0	0	0	0	0
uphold	467	past	Low	2C-6	24	2	0	0	0	0	0	0
creep	455	past	Low	2C-6	24	2	0	0	0	0	0	0
shine	452	past	Low	2C-8	26	2	0	0	0	0	0	0
speed	425	past	Low	2C-6	24	2	0	0	0	0	0	0
rewrite	403	past	Low	3A-1	33	3	0	0	0	0	0	0
mistake	383	past	Low	2B-1	12	2	0	0	0	0	0	0
overtake	377	past	Low	2B-1	12	2	0	0	0	0	0	0
pen	336	past	Low	1C-1	3	1	0	0	0	0	0	0
forecast	323	past	Low	1A-1	1	1	0	0	0	0	0	0
mislead	316	past	Low	2C-6	24	2	0	0	0	0	0	0
string	308	past	Low	2C-1	19	2	0	0	0	0	0	0
fling	307	past	Low	2C-1	19	2	3A-2	34	3	flang	2	increase
undertake	305	past	Low	2B-1	12	2	0	0	0	0	0	0
cling	298	past	Low	2C-1	19	2	0	0	0	0	0	0
shrink	271	past	Low	3A-2	34	3	2C-1	19	2	shrunk	4	decrease
weave	262	past	Low	2D-4	31	2	0	0	0	0	0	0
withhold	261	past	Low	2C-6	24	2	0	0	0	0	0	0
overthrow	258	past	Low	2B-3	14	2	0	0	0	0	0	0
stride	253	past	Low	3A-1	33	3	0	0	0	0	0	0
remake	243	past	Low	1C-5	7	2	0	0	0	0	0	0
plead	236	past	Low	2C-6	24	2	0	0	0	0	0	0
sweat	232	past	Low	1A-1	1	1	0	0	0	0	0	0
outgrow	228	past	Low	2B-3	14	2	0	0	0	0	0	0
kneel	224	past	Low	2C-6	24	2	0	0	0	0	0	0
withstand	198	past	Low	2C-3	21	2	0	0	0	0	0	0
inset	173	past	Low	1A-1	1	1	0	0	0	0	0	0
foresee	172	past	Low	2B-5	16	2	0	0	0	0	0	0
breed	172	past	Low	2C-6	24	2	0	0	0	0	0	0
stink	170	past	Low	3A-2	34	3	2C-1	19	2	stunk	671	decrease
spell	159	past	Low	1C-2	4	2	0	0	0	0	0	0
forbid	159	past	Low	2B-2	13	2	0	0	0	0	0	0
podcast	159	past	Low	1A-1	1	1	0	0	0	0	0	0
bend	146	past	Low	1C-6	8	1	0	0	0	0	0	0
offset	144	past	Low	1A-1	1	1	0	0	0	0	0	0
slit	141	past	Low	1A-1	1	1	0	0	0	0	0	0
recast	139	past	Low	1A-1	1	1	0	0	0	0	0	0
bust	136	past	Low	1A-1	1	1	0	0	0	0	0	0
babysit	128	past	Low	2C-2	20	2	0	0	0	0	0	0
befall	128	past	Low	2B-7	18	2	0	0	0	0	0	0
strive	118	past	Low	3A-1	33	3	0	0	0	0	0	0
rid	116	past	Low	1A-1	1	1	0	0	0	0	0	0
repay	107	past	Low	1C-7	9	1	0	0	0	0	0	0
sling	105	past	Low	2C-1	19	2	0	0	0	0	0	0
foretell	98	past	Low	2C-8	26	2	0	0	0	0	0	0
tread	97	past	Low	2D-5	32	2	0	0	0	0	0	0
outrun	96	past	Low	2A-2	11	2	0	0	0	0	0	0
dwell	89	past	Low	1C-2	4	1	0	0	0	0	0	0
wet	87	past	Low	1A-1	1	1	0	0	0	0	0	0

spoil	85	past	Low	1C-2	4	1	0	0	0	0	0	0
lean	84	past	Low	2C-6	24	2	0	0	0	0	0	0
behold	84	past	Low	2C-6	24	2	0	0	0	0	0	0
override	82	past	Low	3A-1	33	3	0	0	0	0	0	0
retell	80	past	Low	2C-8	26	2	0	0	0	0	0	0
outshoot	80	past	Low	2C-7	25	2	0	0	0	0	0	0
spill	79	past	Low	1C-2	4	1	0	0	0	0	0	0
overrun	79	past	Low	2A-2	11	2	0	0	0	0	0	0
bite	78	past	Low	2D-2	29	2	0	0	0	0	0	0
partake	77	past	Low	2B-1	12	2	0	0	0	0	0	0
cleave	76	past	Low	2D-4	31	2	2C-6	24	2	cleft	5	0
retake	76	past	Low	2B-1	12	2	0	0	0	0	0	0
oversleep	72	past	Low	2C-6	24	2	0	0	0	0	0	0
undercut	71	past	Low	1A-1	1	1	0	0	0	0	0	0
underwrite	66	past	Low	3A-1	33	3	0	0	0	0	0	0
ken	61	past	Low	1C-1	3	1	0	0	0	0	0	0
waylay	58	past	Low	1C-7	9	1	0	0	0	0	0	0
mishear	57	past	Low	2C-4	22	2	0	0	0	0	0	0
slink	57	past	Low	2C-1	19	2	0	0	0	0	0	0
rewind	56	past	Low	2C-5	23	2	0	0	0	0	0	0
overspend	56	past	Low	1C-6	8	1	0	0	0	0	0	0
overshoot	56	past	Low	2C-7	25	2	0	0	0	0	0	0
wring	50	past	Low	2C-1	19	2	3A-2	34	3	wrang	2	increase
inbreed	48	past	Low	2C-6	24	2	0	0	0	0	0	0
smite	46	past	Low	3A-1	33	3	0	0	0	0	0	0
rethink	44	past	Low	2C-9	27	2	0	0	0	0	0	0
beget	41	past	Low	2D-5	32	2	2C-2	20	2	begat	154	0
underthrow	35	past	Low	2B-3	14	2	0	0	0	0	0	0
abide	33	past	Low	2C-8	26	2	0	0	0	0	0	0
outshine	33	past	Low	2C-8	26	2	0	0	0	0	0	0
knit	31	past	Low	1A-1	1	1	0	0	0	0	0	0
unsay	31	past	Low	2C-6	24	2	0	0	0	0	0	0
bespeak	26	past	Low	2D-4	31	2	0	0	0	0	0	0
unwind	23	past	Low	2C-5	23	2	0	0	0	0	0	0
redraw	23	past	Low	2B-4	15	2	0	0	0	0	0	0
unbind	22	past	Low	2C-5	23	2	0	0	0	0	0	0
forsake	22	past	Low	2B-1	12	2	0	0	0	0	0	0
unstick	21	past	Low	2C-1	19	2	0	0	0	0	0	0
burn	20	past	Low	1C-1	3	1	0	0	0	0	0	0
cowrite	19	past	Low	3A-1	33	3	0	0	0	0	0	0
shoe	18	past	Low	2C-7	25	2	0	0	0	0	0	0
miscast	14	past	Low	1A-1	1	1	0	0	0	0	0	0
recut	13	past	Low	1A-1	1	1	0	0	0	0	0	0
unmake	12	past	Low	1C-5	7	1	0	0	0	0	0	0
smell	12	past	Low	1C-2	4	1	0	0	0	0	0	0
foreknow	11	past	Low	2B-3	14	2	0	0	0	0	0	0
typeset	11	past	Low	1A-1	1	1	0	0	0	0	0	0
inlay	11	past	Low	1C-7	9	1	0	0	0	0	0	0
intercut	11	past	Low	1A-1	1	1	0	0	0	0	0	0
betake	11	past	Low	2B-1	12	2	0	0	0	0	0	0
heave	10	past	Low	2D-4	31	2	0	0	0	0	0	0
typecast	9	past	Low	1A-1	1	1	0	0	0	0	0	0
rerun	9	past	Low	2A-2	11	2	0	0	0	0	0	0
overfeed	9	past	Low	2C-6	24	2	0	0	0	0	0	0
uppercut	9	past	Low	1A-1	1	1	0	0	0	0	0	0
thrive	7	past	Low	3A-1	33	3	0	0	0	0	0	0
overwrite	7	past	Low	3A-1	33	3	0	0	0	0	0	0
wed	7	past	Low	1A-1	1	1	0	0	0	0	0	0

sting	7	past	Low	2C-1	19	2	0	0	0	0	0	0
chide	6	past	Low	2D-2	29	2	0	0	0	0	0	0
overblow	5	past	Low	2B-3	14	2	0	0	0	0	0	0
gird	5	past	Low	1C-6	8	1	0	0	0	0	0	0
overdraw	4	past	Low	2B-4	15	2	0	0	0	0	0	0
handwrite	4	past	Low	3A-1	33	3	0	0	0	0	0	0
unfreeze	4	past	Low	2D-4	31	2	0	0	0	0	0	0
bestride	4	past	Low	3A-1	33	3	0	0	0	0	0	0
bless	3	past	Low	1C-4	6	1	0	0	0	0	0	0
bereave	3	past	Low	2C-6	24	2	0	0	0	0	0	0
overhang	3	past	Low	2C-1	19	2	0	0	0	0	0	0
stave	2	past	Low	2C-8	26	2	0	0	0	0	0	0
strip	2	past	Low	1C-3	5	1	0	0	0	0	0	0
clap	2	past	Low	1C-3	5	1	0	0	0	0	0	0
overpay	1465	perfect	Low	1C-7	9	1	0	0	0	0	0	0
thrust	893	perfect	Low	1A-1	1	1	0	0	0	0	0	0
uphold	530	perfect	Low	2C-6	24	2	0	0	0	0	0	0
creep	472	perfect	Low	2C-6	24	2	0	0	0	0	0	0
shine	260	perfect	Low	2C-8	26	2	0	0	0	0	0	0
speed	246	perfect	Low	2C-6	24	2	0	0	0	0	0	0
rewrite	770	perfect	Low	3A-1	33	3	0	0	0	0	0	0
mistake	3914	perfect	Low	2B-1	12	2	0	0	0	0	0	0
overtake	713	perfect	Low	2B-1	12	2	0	0	0	0	0	0
pen	0	perfect	Low	1C-1	3	1	0	0	0	0	0	0
forecast	570	perfect	Low	1A-1	1	1	0	0	0	0	0	0
mislead	508	perfect	Low	2C-6	24	2	0	0	0	0	0	0
string	672	perfect	Low	2C-1	19	2	0	0	0	0	0	0
fling	387	perfect	Low	2C-1	19	2	0	0	0	0	0	0
undertake	788	perfect	Low	2B-1	12	2	0	0	0	0	0	0
cling	122	perfect	Low	2C-1	19	2	0	0	0	0	0	0
shrink	810	perfect	Low	3A-2	34	3	0	0	0	0	0	0
weave	1453	perfect	Low	2D-4	31	2	0	0	0	0	0	0
withhold	644	perfect	Low	2C-6	24	2	0	0	0	0	0	0
overthrow	256	perfect	Low	2B-3	14	2	0	0	0	0	0	0
stride	0	perfect	Low	3A-1	33	3	0	0	0	0	0	0
remake	533	perfect	Low	1C-5	7	1	0	0	0	0	0	0
plead	123	perfect	Low	2C-6	24	2	0	0	0	0	0	0
sweat	0	perfect	Low	1A-1	1	1	0	0	0	0	0	0
outgrow	517	perfect	Low	2B-3	14	2	0	0	0	0	0	0
kneel	44	perfect	Low	2C-6	24	2	0	0	0	0	0	0
withstand	80	perfect	Low	2C-3	21	2	0	0	0	0	0	0
inset	23	perfect	Low	1A-1	1	1	0	0	0	0	0	0
foresee	240	perfect	Low	2B-5	16	2	0	0	0	0	0	0
breed	735	perfect	Low	2C-6	24	2	0	0	0	0	0	0
stink	259	perfect	Low	3A-2	34	3	0	0	0	0	0	0
spell	182	perfect	Low	1C-2	4	1	0	0	0	0	0	0
forbid	1979	perfect	Low	2B-2	13	2	0	0	0	0	0	0
podcast	12	perfect	Low	1A-1	1	1	0	0	0	0	0	0
bend	30	perfect	Low	1C-6	8	1	0	0	0	0	0	0
offset	985	perfect	Low	1A-1	1	1	0	0	0	0	0	0
slit	51	perfect	Low	1A-1	1	1	0	0	0	0	0	0
recast	261	perfect	Low	1A-1	1	1	0	0	0	0	0	0
bust	1	perfect	Low	1A-1	1	1	0	0	0	0	0	0
babysit	39	perfect	Low	2C-2	20	2	0	0	0	0	0	0
befall	102	perfect	Low	2B-7	18	2	0	0	0	0	0	0
strive	33	perfect	Low	3A-1	33	3	0	0	0	0	0	0
rid	479	perfect	Low	1A-1	1	1	0	0	0	0	0	0
repay	1010	perfect	Low	1C-7	9	1	0	0	0	0	0	0

sling	175	perfect	Low	2C-1	19	2	0	0	0	0	0	0
foretell	102	perfect	Low	2C-8	26	2	0	0	0	0	0	0
tread	136	perfect	Low	2D-5	32	2	0	0	0	0	0	0
outrun	44	perfect	Low	2A-2	11	2	0	0	0	0	0	0
dwell	47	perfect	Low	1C-2	4	1	0	0	0	0	0	0
wet	0	perfect	Low	1A-1	1	1	0	0	0	0	0	0
spoil	375	perfect	Low	1C-2	4	1	0	0	0	0	0	0
lean	29	perfect	Low	2C-6	24	2	0	0	0	0	0	0
behold	28	perfect	Low	2C-6	24	2	0	0	0	0	0	0
override	112	perfect	Low	3A-1	33	3	0	0	0	0	0	0
retell	149	perfect	Low	2C-8	26	2	0	0	0	0	0	0
outshoot	59	perfect	Low	2C-7	25	2	0	0	0	0	0	0
spill	172	perfect	Low	1C-2	4	1	0	0	0	0	0	0
overrun	758	perfect	Low	2A-2	11	2	0	0	0	0	0	0
bite	1265	perfect	Low	2D-2	29	2	2D-2	29	2	bit	25	0
partake	78	perfect	Low	2B-1	12	2	0	0	0	0	0	0
cleave	1	perfect	Low	2D-4	31	2	2C-6	24	2	cleft	10	0
retake	25	perfect	Low	2B-1	12	2	0	0	0	0	0	0
oversleep	20	perfect	Low	2C-6	24	2	0	0	0	0	0	0
undercut	111	perfect	Low	1A-1	1	1	0	0	0	0	0	0
underwrite	168	perfect	Low	3A-1	33	3	0	0	0	0	0	0
ken	111	perfect	Low	1C-1	3	1	0	0	0	0	0	0
waylay	106	perfect	Low	1C-7	9	1	0	0	0	0	0	0
mishear	109	perfect	Low	2C-4	22	2	0	0	0	0	0	0
slink	25	perfect	Low	2C-1	19	2	0	0	0	0	0	0
rewind	11	perfect	Low	2C-5	23	2	0	0	0	0	0	0
overspend	196	perfect	Low	1C-6	8	1	0	0	0	0	0	0
overshoot	36	perfect	Low	2C-7	25	2	0	0	0	0	0	0
wring	90	perfect	Low	2C-1	19	2	0	0	0	0	0	0
inbreed	83	perfect	Low	2C-6	24	2	0	0	0	0	0	0
smite	1019	perfect	Low	3A-1	33	3	0	0	0	0	0	0
rethink	68	perfect	Low	2C-9	27	2	0	0	0	0	0	0
beget	149	perfect	Low	2D-5	32	2	0	0	0	0	0	0
underthrow	4	perfect	Low	2B-3	14	2	0	0	0	0	0	0
abide	21	perfect	Low	2C-8	26	2	0	0	0	0	0	0
outshine	42	perfect	Low	2C-8	26	2	0	0	0	0	0	0
knit	3180	perfect	Low	1A-1	1	1	0	0	0	0	0	0
unsay	137	perfect	Low	2C-6	24	2	0	0	0	0	0	0
bespeak	1	perfect	Low	2D-4	31	2	0	0	0	0	0	0
unwind	84	perfect	Low	2C-5	23	2	0	0	0	0	0	0
redraw	0	perfect	Low	2B-4	15	2	0	0	0	0	0	0
unbind	30	perfect	Low	2C-5	23	2	0	0	0	0	0	0
forsake	380	perfect	Low	2B-1	12	2	0	0	0	0	0	0
unstick	78	perfect	Low	2C-1	19	2	0	0	0	0	0	0
burn	1204	perfect	Low	1C-1	3	1	0	0	0	0	0	0
cowrite	6	perfect	Low	3A-1	33	3	0	0	0	0	0	0
shoe	45	perfect	Low	2C-7	25	2	0	0	0	0	0	0
miscast	166	perfect	Low	1A-1	1	1	0	0	0	0	0	0
recut	10	perfect	Low	1A-1	1	1	0	0	0	0	0	0
unmake	31	perfect	Low	1C-5	7	1	0	0	0	0	0	0
smell	260	perfect	Low	1C-2	4	1	0	0	0	0	0	0
foreknow	5	perfect	Low	2B-3	14	2	0	0	0	0	0	0
typeset	39	perfect	Low	1A-1	1	1	0	0	0	0	0	0
inlay	30	perfect	Low	1C-7	9	1	0	0	0	0	0	0
intercut	47	perfect	Low	1A-1	1	1	0	0	0	0	0	0
betake	3	perfect	Low	2B-1	12	2	0	0	0	0	0	0
heave	9	perfect	Low	2D-4	31	2	0	0	0	0	0	0
typecast	131	perfect	Low	1A-1	1	1	0	0	0	0	0	0

rerun	0	perfect	Low	2A-2	11	2	0	0	0	0	0	0
overfeed	31	perfect	Low	2C-6	24	2	0	0	0	0	0	0
uppercut	3	perfect	Low	1A-1	1	1	0	0	0	0	0	0
thrive	0	perfect	Low	3A-1	33	3	0	0	0	0	0	0
overwrite	121	perfect	Low	3A-1	33	3	0	0	0	0	0	0
wed	653	perfect	Low	1A-1	1	1	0	0	0	0	0	0
sting	746	perfect	Low	2C-1	19	2	0	0	0	0	0	0
chide	6	perfect	Low	2D-2	29	2	0	0	0	0	0	0
overblow	4	perfect	Low	2B-3	14	2	0	0	0	0	0	0
gird	7	perfect	Low	1C-6	8	1	0	0	0	0	0	0
overdraw	240	perfect	Low	2B-4	15	2	0	0	0	0	0	0
handwrite	0	perfect	Low	3A-1	33	3	0	0	0	0	0	0
unfreeze	46	perfect	Low	2D-4	31	2	0	0	0	0	0	0
bestride	0	perfect	Low	3A-1	33	3	0	0	0	0	0	0
bless	52	perfect	Low	1C-4	6	1	0	0	0	0	0	0
bereave	0	perfect	Low	2C-6	24	2	0	0	0	0	0	0
overhang	7	perfect	Low	2C-1	19	2	0	0	0	0	0	0
stave	1	perfect	Low	2C-8	26	2	0	0	0	0	0	0
strip	0	perfect	Low	1C-3	5	1	0	0	0	0	0	0
clap	0	perfect	Low	1C-3	5	1	0	0	0	0	0	0
Total	51,121										873	

Appendix 8: Word frequency of IVs and RFs with high frequency split by form and time from WebCorp Corpu

Verbs	Past (1995-2002)	Past (2003-2010)	Perfect (1995-2002)	Perfect (2003-2010)	RFs -past (1995-2002)	RFs -past (2003-2010)	RFs -perfect (1995-2002)	RFs -perfect (2003-2010)
be	22373	3195864	5038	1073733	0	0	0	3
have	5625	1138948	486	103161	0	19	0	3
do	2601	523758	1203	241543	0	0	0	0
say	2504	772020	397	51275	0	7	0	3
get	1771	361548	274	44202	0	0	0	0
make	1014	246913	1044	186546	0	14	0	2
go	867	204894	354	85168	0	3	0	1
think	972	195244	153	31803	0	10	0	2
come	860	184097	287	76546	0	1	0	2
take	697	169528	396	96220	0	7	0	7
tell	452	127733	166	38341	0	0	0	1
see	624	121896	840	169755	0	0	0	0
find	576	121684	393	68096	0	3	0	1
write	503	121154	658	73549	0	4	0	8
give	471	97628	625	114475	0	6	0	4
know	497	96455	528	75156	0	8	0	2
leave	322	86832	298	81828	0	0	0	0
put	351	78483	237	51994	0	30	0	22
feel	253	75281	38	12878	0	6	0	1
win	257	69464	101	30288	0	2	0	1
lose	341	66249	263	62919	0	5	0	1
hit	200	42601	76	27966	0	3	0	0
become	347	52338	183	44311	0	1	0	2

buy	306	51348	153	15812	0	7	0	3
begin	202	47706	44	10024	0	0	0	0
spend	255	46319	156	29828	0	4	0	0
hear	214	42951	394	71452	0	38	0	23
run	180	40129	132	30850	0	1	0	2
bring	200	40093	126	25818	0	0	0	1
keep	182	39055	59	15787	0	1	0	1
set	147	37312	325	64346	0	1	0	2
send	204	28092	119	21912	0	3	0	1
fall	177	30967	59	12112	0	5	0	6
meet	99	29325	104	19848	0	0	0	0
read	244	28600	506	63302	0	9	0	2
mean	147	27878	125	21170	0	3	0	0
throw	54	27597	90	22443	0	3	0	5
lead	107	27213	82	18416	0	0	0	0
break	131	26864	68	16366	0	3	0	2
pay	190	26202	433	56781	0	2	0	2
catch	49	26434	127	6278	0	20	0	11
choose	110	23454	57	16626	0	22	0	2
grow	170	24345	92	19228	0	15	0	20
speak	73	23291	47	10067	0	0	0	0
beat	70	22921	28	9009	0	8	0	5
let	104	22649	30	5809	0	1	0	0
hold	94	22008	216	30052	0	5	0	0
sit	69	21805	9	2961	0	1	0	1
forget	76	20773	79	15596	0	0	0	2
cut	42	19081	44	16773	0	3	0	0
sell	144	18831	302	30572	0	0	0	0
stand	50	16395	33	1308	0	3	0	1
shoot	57	13929	35	8587	0	6	0	0
drive	57	13881	46	10599	0	6	0	4
wear	25	12467	34	8094	0	1	0	0
eat	49	11602	34	6557	0	5	0	4
draw	48	11355	77	16997	0	9	0	3
teach	48	10972	85	9798	0	11	0	5
bet	31	10881	0	307	0	5	0	3
blow	54	10128	50	2784	0	0	0	58
rise	24	10040	20	2765	0	4	0	3
strike	28	8581	28	5496	0	6	0	7
build	70	7912	179	28755	0	6	0	8
shut	28	7767	36	6508	0	0	0	0
stick	52	7612	101	21797	0	20	0	10
wake	44	7531	5	922	0	9	0	6
steal	29	7223	22	7545	0	0	0	0
cost	90	7148	5	2260	0	55	0	84

understand	51	7088	30	6212	0	0	0	1
quit	31	6674	9	1383	0	3	0	3
hang	28	6175	12	3100	0	15	6	399
fly	24	6018	8	2062	1	110	0	8
fight	28	5699	30	3375	0	2	0	0
sing	24	5457	8	2242	0	77	0	125
lie	22	5342	0	180	0	0	0	0
lay	25	5145	42	10244	0	167	0	180
sleep	24	4870	0	4	0	2	0	4
cast	20	4726	33	6021	0	34	0	21
seek	17	3846	19	2979	0	9	0	8
wind	23	3755	1	460	0	0	0	0
dig	19	3689	0	17	0	13	0	8
light	33	3623	17	3062	0	4	0	209
shake	6	3482	4	1636	0	4	0	4
hurt	16	3411	95	21753	0	9	0	1
split	15	3384	23	3827	0	0	0	1
spread	15	3358	22	3991	0	2	0	6
ride	12	3168	0	861	0	1	0	0
drink	27	3075	0	1	0	1	0	0
tear	11	2759	20	5302	0	0	0	0
upset	9	2704	49	12662	0	0	0	0
ring	5	2269	4	1004	0	0	0	0
shed	2	2197	4	553	0	8	0	2
shit	9	2137	5	1611	0	2	0	5
deal	7	2095	30	5840	0	3	0	18
sweep	4	2074	13	3089	0	1	0	1
bear	8	2045	167	30882	0	5	0	5
hide	7	1721	82	6197	0	3	0	0
swing	3	1672	6	665	0	1	0	3
slide	2	1560	0	175	0	6	0	3
freeze	5	1542	13	3122	0	2	0	6
swear	8	1519	9	2177	0	1	0	1
arise	13	1492	2	382	0	2	0	3
feed	8	1461	26	1461	0	0	0	1
learn	2	1303	1	1413	119	31476	75	24292
undergo	6	1289	0	814	0	0	0	0
withdraw	1	1292	5	1645	0	3	0	0
rebuild	5	1260	0	0	0	0	0	0
spin	3	1235	6	1248	0	8	0	3
sink	4	1204	8	1249	0	3	0	1
flee	6	1097	3	334	0	0	0	0
sneak	6	1088	1	382	3	482	0	1
lend	2	1041	1	685	0	6	0	4
spit	4	1017	1	317	0	2	0	2

awake	5	891	2	215	0	4	0	1
bid	3	853	0	166	0	1	0	2
dream	2	837	0	35	6	1916	4	1506
broadcast	5	752	12	2506	0	1506	0	134
burst	7	733	0	0	0	12	0	7
swim	5	670	0	52	0	0	0	0
weep	3	631	0	0	0	7	0	2
spring	7	596	11	1117	0	0	0	1
bleed	4	585	0	6	0	0	0	0
grind	4	575	3	586	0	73	0	8
leap	4	540	0	3	1	498	0	67
Total	49,951	9,310,000	19,671	3,823,375	130	36,914	85	27,410
	9,359,951		3,843,046		37,044		27,495	
	13,202,997				64,539			
13,267,536								

Appendix 9: Word frequency of IVs and RFs with low frequency split by form and time from WebCorp Corpus

Verbs	Past (1995-2002)	Past (2003-2010)	Perfect (1995-2002)	Perfect (2003-2010)	RFs --past (1995-2002)	RFs -past (2003-2010)	RFs -perfect (1995-2002)	RFs -perfect (2003-2010)
forgive	9	497	17	2629	0	2	0	2
overpay	0	496	2	1461	0	9	0	14
thrust	1	472	5	886	0	10	0	8
uphold	2	465	2	528	0	0	0	0
creep	2	452	2	470	11	222	6	421
shine	4	448	4	256	1	441	1	269
speed	2	423	0	245	0	75	0	5
rewrite	4	392	5	765	0	0	0	0
mistake	2	380	30	3874	0	2	0	1
overtake	0	377	5	708	0	0	0	1
pen	1	334	0	0	0	174	12	2031
forecast	0	323	1	569	0	55	0	205
mislead	1	315	8	500	0	0	0	1
string	2	306	3	670	0	0	0	0
fling	2	305	6	380	0	0	0	0
undertake	0	305	6	781	0	0	0	0
cling	0	298	0	122	0	4	0	0
shrink	1	268	4	805	0	2	0	0
weave	2	260	7	1445	0	193	0	126
withhold	6	255	6	636	0	0	0	0
overthrow	0	254	1	253	0	0	0	0
stride	0	253	0	0	0	2	0	1
remake	1	242	5	528	0	0	0	2
plead	3	233	0	123	5	1503	2	280

sweat	3	229	0	0	0	12	1	248
outgrow	1	227	2	515	0	0	0	0
oudo	1	223	5	926	0	0	0	0
kneel	0	223	0	44	0	41	0	18
withstand	0	192	0	80	0	0	0	0
redo	2	183	3	512	0	0	0	0
inset	0	173	0	23	0	0	0	0
breed	1	171	1	734	0	0	0	0
undo	2	169	5	1259	0	0	0	0
stink	1	169	1	258	0	1	0	0
spell	0	159	1	181	0	92	37	3051
forbid	0	158	18	1960	0	0	0	0
podcast	0	159	0	12	0	7	0	8
bend	27	119	0	30	0	15	0	5
offset	1	143	5	979	0	0	0	0
slit	0	141	1	50	0	1	0	1
recast	0	139	0	261	0	7	0	15
bust	2	134	0	1	14	1287	6	1314
babysit	1	128	0	39	0	2	0	0
befall	0	128	1	101	0	0	0	1
strive	0	118	1	32	0	62	0	107
rid	0	116	2	476	0	4	0	3
repay	0	107	13	997	0	4	0	7
sling	1	104	2	173	0	0	0	2
foretell	0	98	1	100	0	0	0	0
tread	1	96	0	136	0	42	0	9
outrun	0	96	0	44	0	0	0	0
dwell	0	89	0	47	0	46	0	25
wet	0	87	0	0	0	6	0	20
spoil	0	85	0	375	3	1307	0	0
lean	0	84	1	28	3	1412	1	446
behold	0	84	0	28	0	0	0	0
override	0	82	1	110	0	0	0	0
retell	1	79	2	147	0	0	0	0
outshoot	0	80	0	59	0	0	0	0
spill	0	78	0	172	4	1142	4	699
overrun	0	79	5	749	0	0	0	0
bite	0	78	5	1257	0	1	0	0
partake	1	76	2	76	0	0	0	0
cleave	0	76	0	1	0	7	1	25
retake	1	75	1	24	0	0	0	0
oversleep	0	72	0	20	0	0	0	0
undercut	1	69	0	111	0	0	0	0
underwrite	0	66	0	168	0	0	0	0
ken	0	61	0	111	0	0	0	2

waylay	0	58	0	106	0	0	0	1
mishear	1	56	1	51	0	0	0	0
slink	0	57	0	25	0	46	0	10
rewind	1	55	0	11	0	5	0	0
overspend	0	56	0	196	0	0	0	0
overshoot	0	56	0	36	0	0	0	0
wring	0	50	0	90	0	0	0	0
inbreed	0	48	0	83	0	0	0	0
smite	0	46	3	1016	0	9	0	12
rethink	2	42	0	68	0	0	0	0
beget	0	41	0	149	0	0	0	0
underthrow	0	35	0	4	0	0	0	0
abide	0	33	0	21	0	22	2	17
outshine	0	33	0	42	0	40	1	27
knit	0	31	10	3170	0	33	1	1188
unsay	0	31	0	136	0	0	0	0
bespeak	1	25	0	1	0	0	0	0
unwind	0	23	0	84	0	0	0	0
redraw	0	23	0	0	0	0	0	0
unbind	0	22	0	30	0	0	0	0
forsake	0	22	2	378	0	2	0	0
unstick	2	19	0	78	0	0	0	0
burn	0	20	6	1196	11	2311	80	5292
cowrite	0	19	0	6	0	0	0	0
shoe	0	18	1	44	0	8	0	14
miscast	0	14	0	166	0	1	0	0
forgo	0	14	0	143	0	2	0	0
recut	0	13	0	10	0	0	0	0
unmake	0	12	0	31	0	0	0	0
smell	0	12	4	256	17	1724	0	4
foreknow	0	11	0	5	0	0	0	0
typeset	0	11	2	37	0	0	0	0
inlay	0	11	0	30	0	1	0	3
intercut	0	11	5	42	0	0	0	0
betake	0	11	0	3	0	0	0	0
dare	0	10	0	0	4	1509	0	9
heave	0	10	0	9	0	191	0	49
typecast	0	9	2	129	0	3	0	5
rerun	0	9	0	0	0	0	0	0
overfeed	1	8	0	31	0	0	0	0
uppercut	0	9	0	3	0	1	0	0
thrive	0	7	0	0	3	758	0	0
overwrite	0	7	0	121	0	0	0	1
sting	0	7	4	744	0	0	0	0
wed	0	7	4	649	0	0	2	311

chide	0	6	0	0	0	174	0	85
overblow	0	5	0	4	0	0	0	0
gird	0	5	0	7	0	39	0	2
overdraw	0	4	0	240	0	0	0	0
handwrite	0	4	0	0	0	0	0	0
unfreeze	0	4	0	46	0	0	0	0
bestride	0	4	0	0	0	0	0	0
bless	0	3	0	52	0	56	30	17375
bereave	0	3	0	0	0	10	0	103
overhang	0	3	0	7	0	1	0	0
Total	101	15018	242	41826	76	15138	187	33881
	15119		42068		15214		34068	
	57187				49282			
	106469							

Appendix 10: Word frequencies of IVs and IFs with high frequency split by form and time from WebCorp Corpus

Verbs	Form	1995-2002	2003-2010	Class	Class-N.	Vowel-N.	Class	Class-change N.	Vowel-change N.	IFs verb	1995-2002	2003-2010	direction of vowel change
say	Past	2504	772020	2C-6	24	2	_	_	_	_	_	_	_
get	Past	1771	361548	2D-5	32	2	_	_	_	_	_	_	_
make	Past	1014	246913	1C-5	7	1	_	_	_	_	_	_	_
think	Past	972	195244	2C-9	27	2	_	_	_	_	_	_	_
come	Past	860	184097	2A-1	10	2	_	_	_	_	_	_	_
take	Past	697	169528	2B-1	12	2	_	_	_	_	_	_	_
tell	Past	452	127733	2C-8	26	2	_	_	_	_	_	_	_
see	Past	624	121896	2B-5	16	2	_	_	_	_	_	_	_
find	Past	576	121684	2C-5	23	2	_	_	_	_	_	_	_
write	Past	503	121154	3A-1	33	3	_	_	_	_	_	_	_
give	Past	471	97628	2B-2	13	2	_	_	_	_	_	_	_
know	Past	497	96455	2B-3	14	2	_	_	_	_	_	_	_
leave	Past	322	86832	2C-6	24	2	_	_	_	_	_	_	_
put	Past	351	78483	1A-1	1	1	_	_	_	_	_	_	_
feel	Past	253	75281	2C-6	24	2	_	_	_	_	_	_	_
win	Past	257	69464	2C-1	19	2	_	_	_	_	_	_	_
lose	Past	341	66249	2C-7	25	2	_	_	_	_	_	_	_
hit	Past	200	42601	1A-1	1	1	_	_	_	_	_	_	_
become	Past	347	52338	2A-1	10	2	_	_	_	_	_	_	_
buy	Past	306	51348	2C-9	27	2	_	_	_	_	_	_	_
begin	Past	202	47706	3A-2	34	3	_	_	_	_	_	_	_
spend	Past	255	46319	1C-6	8	1	_	_	_	_	_	_	_
hear	Past	214	42951	2C-4	22	2	_	_	_	_	_	_	_
run	Past	180	40129	2A-2	11	2	_	_	_	_	_	_	_
bring	Past	200	40093	2C-9	27	2	2C-1	19	2	brung	_	27	_
keep	Past	182	39055	2C-6	24	2	_	_	_	_	_	_	_
set	Past	147	37312	1A-1	1	1	_	_	_	_	_	_	_
send	Past	204	28092	1C-6	8	1	_	_	_	_	_	_	_
fall	Past	177	30967	2B-7	18	2	_	_	_	_	_	_	_
meet	Past	99	29325	2C-6	24	2	_	_	_	_	_	_	_
read	Past	244	28600	2C-6	24	2	_	_	_	_	_	_	_

mean	Past	147	27878	2C-6	24	2	_	_	_	_	_	_	_
throw	Past	54	27597	2B-3	14	2	_	_	_	_	_	_	_
lead	Past	107	27213	2C-6	24	2	_	_	_	_	_	_	_
break	Past	131	26864	2D-4	31	2	_	_	_	_	_	_	_
pay	Past	190	26202	1C-7	9	1	_	_	_	_	_	_	_
catch	Past	49	26434	2C-9	27	2	_	_	_	_	_	_	_
choose	Past	110	23454	2D-4	31	2	_	_	_	_	_	_	_
grow	Past	170	24345	2B-3	14	2	_	_	_	_	_	_	_
speak	Past	73	23291	2D-4	31	2	_	_	_	_	_	_	_
beat	Past	70	22921	1B-1	2	1	_	_	_	_	_	_	_
let	Past	104	22649	1A-1	1	1	_	_	_	_	_	_	_
hold	Past	94	22008	2C-6	24	2	_	_	_	_	_	_	_
sit	Past	69	21805	2C-2	20	2	_	_	_	_	_	_	_
forget	Past	76	20773	2D-5	32	2	_	_	_	_	_	_	_
cut	Past	42	19081	1A-1	1	1	_	_	_	_	_	_	_
sell	Past	144	18831	2C-8	26	2	_	_	_	_	_	_	_
stand	Past	50	16395	2C-3	21	2	_	_	_	_	_	_	_
shoot	Past	57	13929	2C-7	25	2	_	_	_	_	_	_	_
drive	Past	57	13881	3A-1	33	3	_	_	_	_	_	_	_
wear	Past	25	12467	2D-1	28	2	_	_	_	_	_	_	_
eat	Past	49	11602	2B-6	17	2	_	_	_	_	_	_	_
draw	Past	48	11355	2B-4	15	2	_	_	_	_	_	_	_
teach	Past	48	10972	2C-9	27	2	_	_	_	_	_	_	_
bet	Past	31	10881	1A-1	1	1	_	_	_	_	_	_	_
blow	Past	54	10128	2B-3	14	2	_	_	_	_	_	_	_
rise	Past	24	10040	3A-1	33	3	_	_	_	_	_	_	_
strike	Past	28	8581	2C-1	19	2	_	_	_	_	_	_	_
build	Past	70	7912	1C-6	8	1	_	_	_	_	_	_	_
shut	Past	28	7767	1A-1	1	1	_	_	_	_	_	_	_
stick	Past	52	7612	2C-1	19	2	_	_	_	_	_	_	_
wake	Past	44	7531	2D-4	31	2	_	_	_	_	_	_	_
steal	Past	29	7223	2D-4	31	2	_	_	_	_	_	_	_
cost	Past	90	7148	1A-1	1	1	_	_	_	_	_	_	_
quit	Past	31	6674	1A-1	1	1	_	_	_	_	_	_	_
understa--nd	Past	51	7088	2C-3	21	2	_	_	_	_	_	_	_
hang	Past	28	6175	2C-1	19	3	_	_	_	_	_	_	_
fly	Past	24	6018	3A-3	35	2	_	_	_	_	_	_	_
fight	Past	28	5699	2C-9	27	3	_	_	_	_	_	_	_
sing	Past	24	5457	3A-2	34	2	2C-1	19	2	sung	_	3	decrease
lie	Past	22	5342	2D-3	30	1	_	_	_	_	_	_	_
lay	Past	25	5145	1C-7	9	2	_	_	_	_	_	_	_
sleep	Past	24	4870	2C-6	24	1	_	_	_	_	_	_	_
cast	Past	20	4726	1A-1	1	3	_	_	_	_	_	_	_
seek	Past	17	3846	2C-9	27	2	_	_	_	_	_	_	_
wind	Past	23	3755	2C-5	23	2	_	_	_	_	_	_	_
dig	Past	19	3689	2C-1	19	2	_	_	_	_	_	_	_
light	Past	33	3623	2D-2	29	2	_	_	_	_	_	_	_
shake	Past	6	3482	2B-1	12	1	_	_	_	_	_	_	_
hurt	Past	16	3411	1A-1	1	1	_	_	_	_	_	_	_
split	Past	15	3384	1A-1	1	1	_	_	_	_	_	_	_
spread	Past	15	3358	1A-1	1	3	_	_	_	_	_	_	_
ride	Past	12	3168	3A-1	33	3	_	_	_	_	_	_	_
drink	Past	27	3075	3A-2	34	2	_	_	_	_	_	_	_
tear	Past	11	2759	2D-1	28	1	_	_	_	_	_	_	_
upset	Past	9	2704	1A-1	1	3	_	_	_	_	_	_	_
ring	Past	5	2269	3A-2	34	1	2C-1	19	2	rung	_	2	decrease
shed	Past	2	2197	1A-1	1	1	_	_	_	_	_	_	_
shit	Past	9	2137	1A-1	1	2	2C-2	20	2	shat	3	237	increase

deal	Past	7	2095	2C-6	24	2	_	_	_	_	_	_	_
sweep	Past	4	2074	2C-6	24	2	_	_	_	_	_	_	_
bear	Past	8	2045	2D-1	28	2	_	_	_	_	_	_	_
hide	Past	7	1721	2D-2	29	2	_	_	_	_	_	_	_
swing	Past	3	1672	2C-1	19	2	3A-2	34	3	swang	_	1	increase
slide	Past	2	1560	2D-2	29	2	_	_	_	_	_	_	_
freeze	Past	5	1542	2D-4	31	2	_	_	_	_	_	_	_
swear	Past	8	1519	2D-1	28	2	_	_	_	_	_	_	_
overcome	Past	0	268	2A-1	10	3	_	_	_	_	_	_	_
arise	Past	13	1492	3A-1	33	2	_	_	_	_	_	_	_
feed	Past	8	1461	2C-6	24	1	_	_	_	_	_	_	_
learn	Past	2	1303	1C-1	3	2	_	_	_	_	_	_	_
withdraw	Past	1	1292	2B-4	15	1	_	_	_	_	_	_	_
rebuild	Past	5	1260	1C-6	8	2	_	_	_	_	_	_	_
spin	Past	3	1235	2C-1	19	3	_	_	_	_	_	_	_
sink	Past	4	1204	3A-2	34	2	2C-1	19	2	sunk	_	587	decrease
flee	Past	6	1097	2C-6	24	2	_	_	_	_	_	_	_
sneak	Past	6	1088	2C-1	19	1	_	_	_	_	_	_	_
lend	Past	2	1041	1C-6	8	1	_	_	_	_	_	_	_
spit	Past	4	1017	1A-1	1	2	2C-2	20	2	spat	1	394	increase
awake	Past	5	891	2D-4	31	1	_	_	_	_	_	_	_
bid	Past	3	853	1A-1	1	2	_	_	_	_	_	_	_
dream	Past	2	837	2C-6	24	1	_	_	_	_	_	_	_
broad-cast	Past	5	752	1A-1	1	1	_	_	_	_	_	_	_
burst	Past	7	733	1A-1	1	2	_	_	_	_	_	_	_
oversee	Past	0	47	2B-5	16	2		_	_	_	_	_	_
swim	Past	5	670	3A-2	34	3	_	_	_	_	_	_	_
weep	Past	3	631	2C-6	24	2	_	_	_	_	_	_	_
spring	Past	7	596	3A-2	34	3	2C-1	19	2	sprung	_	3	decrease
bleed	Past	4	585	2C-6	24	2	_	_	_	_	_	_	_
grind	Past	4	575	2C-5	23	2	_	_	_	_	_	_	_
leap	Past	4	540	2C-6	24	2	_	_	_	_	_	_	_
forgive	Past	0	277	2B-2	13	2	_	_	_	_	_	_	_
say	Perfect	397	51275	2C-6	24	2	_	_	_	_	_	_	_
get	Perfect	274	44202	2D-5	32	2	_	_	_	_	_	_	_
make	Perfect	1044	186546	1C-5	7	1	_	_	_	_	_	_	_
think	Perfect	153	31803	2C-9	27	2	_	_	_	_	_	_	_
come	Perfect	287	76546	2A-1	10	2	_	_	_	_	_	_	_
take	Perfect	396	96220	2B-1	12	2	_	_	_	_	_	_	_
tell	Perfect	166	38341	2C-8	26	2	_	_	_	_	_	_	_
see	Perfect	840	169755	2B-5	16	2	_	_	_	_	_	_	_
find	Perfect	393	68096	2C-5	23	2	_	_	_	_	_	_	_
write	Perfect	658	73549	3A-1	33	3	_	_	_	_	_	_	_
give	Perfect	625	114475	2B-2	13	2	_	_	_	_	_	_	_
know	Perfect	528	75156	2B-3	14	2	_	_	_	_	_	_	_
leave	Perfect	298	81828	2C-6	24	2	_	_	_	_	_	_	_
put	Perfect	237	51994	1A-1	1	1	_	_	_	_	_	_	_
feel	Perfect	38	12878	2C-6	24	2	_	_	_	_	_	_	_
win	Perfect	101	30288	2C-1	19	2	_	_	_	_	_	_	_
lose	Perfect	263	62919	2C-7	25	2	_	_	_	_	_	_	_
hit	Perfect	76	27966	1A-1	1	1	_	_	_	_	_	_	_
become	Perfect	183	44311	2A-1	10	2	_	_	_	_	_	_	_
buy	Perfect	153	15812	2C-9	27	2	_	_	_	_	_	_	_
begin	Perfect	44	10024	3A-2	34	3	_	_	_	_	_	_	_
spend	Perfect	156	29828	1C-6	8	1	_	_	_	_	_	_	_
hear	Perfect	394	71452	2C-4	22	2	_	_	_	_	_	_	_
run	Perfect	132	30850	2A-2	11	2	_	_	_	_	_	_	_
bring	Perfect	126	25818	2C-9	27	2	2C-1	19	2	brung	_	5	_

keep	Perfect	59	15787	2C-6	24	2	_	_	_	_	_	_	_
set	Perfect	325	64346	1A-1	1	1	_	_	_	_	_	_	_
send	Perfect	119	21912	1C-6	8	1	_	_	_	_	_	_	_
fall	Perfect	59	12112	2B-7	18	2	_	_	_	_	_	_	_
meet	Perfect	104	19848	2C-6	24	2	_	_	_	_	_	_	_
read	Perfect	506	63302	2C-6	24	2	_	_	_	_	_	_	_
mean	Perfect	125	21170	2C-6	24	2	_	_	_	_	_	_	_
throw	Perfect	90	22443	2B-3	14	2	_	_	_	_	_	_	_
lead	Perfect	82	18416	2C-6	24	2	_	_	_	_	_	_	_
break	Perfect	68	16366	2D-4	31	2	_	_	_	_	_	_	_
pay	Perfect	433	56781	1C-7	9	1	_	_	_	_	_	_	_
catch	Perfect	127	6278	2C-9	27	2	_	_	_	_	_	_	_
choose	Perfect	57	16626	2D-4	31	2	_	_	_	_	_	_	_
grow	Perfect	92	19228	2B-3	14	2	_	_	_	_	_	_	_
speak	Perfect	47	10067	2D-4	31	2	_	_	_	_	_	_	_
beat	Perfect	28	9009	1B-1	2	1	_	_	_	_	_	_	_
let	Perfect	30	5809	1A-1	1	1	_	_	_	_	_	_	_
hold	Perfect	216	30052	2C-6	24	2	_	_	_	_	_	_	_
sit	Perfect	9	2961	2C-2	20	2	_	_	_	_	_	_	_
forget	Perfect	79	15596	2D-5	32	2	2C-7	25	2	forgot	8	2786	decrease
cut	Perfect	44	16773	1A-1	1	1	_	_	_	_	_	_	_
sell	Perfect	302	30572	2C-8	26	2	_	_	_	_	_	_	_
stand	Perfect	33	1308	2C-3	21	2	_	_	_	_	_	_	_
shoot	Perfect	35	8587	2C-7	25	2	_	_	_	_	_	_	_
drive	Perfect	46	10599	3A-1	33	3	_	_	_	_	_	_	_
wear	Perfect	34	8094	2D-1	28	2	_	_	_	_	_	_	_
eat	Perfect	34	6557	2B-6	17	2	_	_	_	_	_	_	_
draw	Perfect	77	16997	2B-4	15	2	_	_	_	_	_	_	_
teach	Perfect	85	9798	2C-9	27	2	_	_	_	_	_	_	_
bet	Perfect	0	307	1A-1	1	1	_	_	_	_	_	_	_
blow	Perfect	50	2784	2B-3	14	2	_	_	_	_	_	_	_
rise	Perfect	20	2765	3A-1	33	3	_	_	_	_	_	_	_
strike	Perfect	28	5496	2C-1	19	2	3A-1	1	3	stricken	5	539	increase
build	Perfect	179	28755	1C-6	8	1	_	_	_	_	_	_	_
shut	Perfect	36	6508	1A-1	1	1	_	_	_	_	_	_	_
stick	Perfect	101	21797	2C-1	19	2	_	_	_	_	_	_	_
wake	Perfect	5	922	2D-4	31	2	_	_	_	_	_	_	_
steal	Perfect	22	7545	2D-4	31	2	_	_	_	_	_	_	_
cost	Perfect	5	2260	1A-1	1	1	_	_	_	_	_	_	_
quit	Perfect	9	1383	1A-1	1	1	_	_	_	_	_	_	_
understand	Perfect	30	6212	2C-3	21	2	_	_	_	_	_	_	_
hang	Perfect	12	3100	2C-1	19	3	_	_	_	_	_	_	_
fly	Perfect	8	2062	3A-3	35	2	_	_	_	_	_	_	_
fight	Perfect	30	3375	2C-9	27	3	_	_	_	_	_	_	_
sing	Perfect	8	2242	3A-2	34	2	_	_	_	_	_	_	_
lie	Perfect	0	180	2D-3	30	1	_	_	_	_	_	_	_
lay	Perfect	42	10244	1C-7	9	2	_	_	_	_	_	_	_
sleep	Perfect	0	4	2C-6	24	1	_	_	_	_	_	_	_
cast	Perfect	33	6021	1A-1	1	3	_	_	_	_	_	_	_
seek	Perfect	19	2979	2C-9	27	2	_	_	_	_	_	_	_
wind	Perfect	1	460	2C-5	23	2	_	_	_	_	_	_	_
dig	Perfect	0	17	2C-1	19	2	_	_	_	_	_	_	_
light	Perfect	17	3062	2D-2	29	2	_	_	_	_	_	_	_
shake	Perfect	4	1636	2B-1	12	1	_	_	_	_	_	_	_
hurt	Perfect	95	21753	1A-1	1	1	_	_	_	_	_	_	_
split	Perfect	23	3827	1A-1	1	1	_	_	_	_	_	_	_
spread	Perfect	22	3991	1A-1	1	3	_	_	_	_	_	_	_
ride	Perfect	0	861	3A-1	33	3	_	_	_	_	_	_	_

drink	Perfect	0	1	3A-2	34	2	_	_	_	_	_	_	_
tear	Perfect	20	5302	2D-1	28	1	_	_	_	_	_	_	_
upset	Perfect	49	12662	1A-1	1	3	_	_	_	_	_	_	_
ring	Perfect	4	1004	3A-2	33	1	_	_	_	_	_	_	_
shed	Perfect	4	553	1A-1	1	1	_	_	_	_	_	_	_
shit	Perfect	5	1611	1A-1	1	2	_	_	_	_	_	_	_
deal	Perfect	30	5840	2C-6	24	2	_	_	_	_	_	_	_
sweep	Perfect	13	3089	2C-6	24	2	_	_	_	_	_	_	_
bear	Perfect	167	30882	2D-1	28	2	_	_	_	_	_	_	_
hide	Perfect	82	6197	2D-2	29	2	_	_	_	_	_	_	_
swing	Perfect	6	665	2C-1	19	2	_	_	_	_	_	_	_
slide	Perfect	0	175	2D-2	29	2	_	_	_	_	_	_	_
freeze	Perfect	13	3122	2D-4	31	2	_	_	_	_	_	_	_
swear	Perfect	9	2177	2D-1	28	2	_	_	_	_	_	_	_
overcome	Perfect	0	787	2A-1	10	3	_	_	_	_	_	_	_
arise	Perfect	2	382	3A-1	33	2	_	_	_	_	_	_	_
feed	Perfect	26	1461	2C-6	24	1	_	_	_	_	_	_	_
learn	Perfect	1	1413	1C-1	3	2	_	_	_	_	_	_	_
withdraw	Perfect	5	1645	2B-4	15	1	_	_	_	_	_	_	_
rebuild	Perfect	0	0	1C-6	8	2	_	_	_	_	_	_	_
spin	Perfect	6	1248	2C-1	19	3	_	_	_	_	_	_	_
sink	Perfect	8	1249	3A-2	34	2	_	_	_	_	_	_	_
flee	Perfect	3	334	2C-6	24	2	_	_	_	_	_	_	_
sneak	Perfect	1	382	2C-1	19	1	_	_	_	_	_	_	_
lend	Perfect	1	685	1C-6	8	1	_	_	_	_	_	_	_
spit	Perfect	1	317	1A-1	1	2	2C-2	20	2	spat	_	142	increase
awake	Perfect	2	215	2D-4	31	1	_	_	_	_	_	_	_
bid	Perfect	0	166	1A-1	1	2	_	_	_	_	_	_	_
dream	Perfect	0	35	2C-6	24	1	_	_	_	_	_	_	_
broadcast	Perfect	12	2506	1A-1	1	1	_	_	_	_	_	_	_
burst	Perfect	0	263	1A-1	1	2	_	_	_	_	_	_	_
oversee	Perfect	0	47	2B-5	16	3	_	_	_	_	_	_	_
swim	Perfect	0	52	3A-2	34	2	_	_	_	_	_	_	_
weep	Perfect	0	0	2C-6	24	3	_	_	_	_	_	_	_
spring	Perfect	11	1117	3A-2	34	2	_	_	_	_	_	_	_
bleed	Perfect	0	6	2C-6	24	2	_	_	_	_	_	_	_
grind	Perfect	3	586	2C-5	23	2	_	_	_	_	_	_	_
leap	Perfect	0	3	2C-6	24	2	_	_	_	_	_	_	_
forgive	Perfect	0	1184	2B-2	13	2	_	_	_	_	_	_	_

Appendix 11: Word frequencies of IVs and IFs with low frequency split by form and time from WebCorp Corpus

verbs	Form	1995-2002	2003-2010	Class	Class-N.	Vowel-N.	Class	Class-change N.	Vowel-change N.	IFs verb	1995-2002	2003-2010	direction of vowel change
overpay	Past	0	496	1C-7	9	1	_	_	_	_	_	_	_
thrust	Past	1	472	1A-1	1	1	_	_	_	_	_	_	_
uphold	Past	2	465	2C-6	24	2	_	_	_	_	_	_	_
creep	Past	2	452	2C-6	24	2	_	_	_	_	_	_	_
shine	Past	4	448	2C-8	26	2	_	_	_	_	_	_	_
speed	Past	2	423	2C-6	24	2	_	_	_	_	_	_	_
rewrite	Past	4	392	3A-1	33	3	_	_	_	_	_	_	_
mistake	Past	2	380	2B-1	12	2	_	_	_	_	_	_	_
overtake	Past	0	377	2B-1	12	2	_	_	_	_	_	_	_
pen	Past	1	334	1C-1	3	1	_	_	_	_	_	_	_
forecast	Past	0	323	1A-1	1	1	_	_	_	_	_	_	_
mislead	Past	1	315	2C-6	24	2	_	_	_	_	_	_	_
string	Past	2	306	2C-1	19	2	_	_	_	_	_	_	_
fling	Past	2	305	2C-1	19	2	3A-2	34	3	flang	_	2	increase
undertake	Past	0	305	2B-1	12	2	_	_	_	_	_	_	_
cling	Past	0	298	2C-1	19	2	_	_	_	_	_	_	_
shrink	Past	1	268	3A-2	34	3	2C-1	19	2	shrunk	_	4	decrease
weave	Past	2	260	2D-4	31	2	_	_	_	_	_	_	_
withhold	Past	6	255	2C-6	24	2	_	_	_	_	_	_	_
overthrow	Past	0	254	2B-3	14	2	_	_	_	_	_	_	_
stride	Past	0	253	3A-1	33	3	_	_	_	_	_	_	_
remake	Past	1	242	1C-5	7	2	_	_	_	_	_	_	_
plead	Past	3	233	2C-6	24	2	_	_	_	_	_	_	_
sweat	Past	3	229	1A-1	1	1	_	_	_	_	_	_	_
outgrow	Past	1	227	2B-3	14	2	_	_	_	_	_	_	_
kneel	Past	0	223	2C-6	24	2	_	_	_	_	_	_	_
withstand	Past	0	192	2C-3	21	2	_	_	_	_	_	_	_
inset	Past	0	173	1A-1	1	1	_	_	_	_	_	_	_
foresee	Past	1	172	2B-5	16	2	_	_	_	_	_	_	_
breed	Past	1	171	2C-6	24	2	_	_	_	_	_	_	_
stink	Past	1	169	3A-2	34	3	2C-1	19	2	stunk	1	670	decrease
spell	Past	0	159	1C-2	4	2	_	_	_	_	_	_	_
forbid	Past	0	158	2B-2	13	2	_	_	_	_	_	_	_
podcast	Past	0	159	1A-1	1	1	_	_	_	_	_	_	_
bend	Past	27	119	1C-6	8	1	_	_	_	_	_	_	_
offset	Past	1	143	1A-1	1	1	_	_	_	_	_	_	_
slit	Past	0	141	1A-1	1	1	_	_	_	_	_	_	_
recast	Past	0	139	1A-1	1	1	_	_	_	_	_	_	_
bust	Past	2	134	1A-1	1	1	_	_	_	_	_	_	_
babysit	Past	1	128	2C-2	20	2	_	_	_	_	_	_	_
befall	Past	0	128	2B-7	18	2	_	_	_	_	_	_	_
strive	Past	0	118	3A-1	33	3	_	_	_	_	_	_	_
rid	Past	0	116	1A-1	1	1	_	_	_	_	_	_	_
repay	Past	0	107	1C-7	9	1	_	_	_	_	_	_	_
sling	Past	1	104	2C-1	19	2	_	_	_	_	_	_	_
foretell	Past	0	98	2C-8	26	2	_	_	_	_	_	_	_
tread	Past	1	96	2D-5	32	2	_	_	_	_	_	_	_
outrun	Past	0	96	2A-2	11	2	_	_	_	_	_	_	_
dwell	Past	0	89	1C-2	4	1	_	_	_	_	_	_	_

wet	Past	0	87	1A-1	1	1	_	_	_	_	_	_	_
spoil	Past	0	85	1C-2	4	1	_	_	_	_	_	_	_
lean	Past	0	84	2C-6	24	2	_	_	_	_	_	_	_
behold	Past	0	84	2C-6	24	2	_	_	_	_	_	_	_
override	Past	0	82	3A-1	33	3	_	_	_	_	_	_	_
retell	Past	1	79	2C-8	26	2	_	_	_	_	_	_	_
outshoot	Past	0	80	2C-7	25	2	_	_	_	_	_	_	_
spill	Past	0	78	1C-2	4	1	_	_	_	_	_	_	_
overrun	Past	0	79	2A-2	11	2	_	_	_	_	_	_	_
bite	Past	0	78	2D-2	29	2	_	_	_	_	_	_	_
partake	Past	1	76	2B-1	12	2	_	_	_	_	_	_	_
cleave	Past	0	76	2D-4	31	2	2C-6	24	2	cleft	_	5	decrease
retake	Past	1	75	2B-1	12	2	_	_	_	_	_	_	_
oversleep	Past	0	72	2C-6	24	2	_	_	_	_	_	_	_
undercut	Past	1	69	1A-1	1	1	_	_	_	_	_	_	_
underwrite	Past	0	66	3A-1	33	3	_	_	_	_	_	_	_
ken	Past	0	61	1C-1	3	1	_	_	_	_	_	_	_
waylay	Past	0	58	1C-7	9	1	_	_	_	_	_	_	_
mishear	Past	1	56	2C-4	22	2	_	_	_	_	_	_	_
slink	Past	0	57	2C-1	19	2	_	_	_	_	_	_	_
rewind	Past	1	55	2C-5	23	2	_	_	_	_	_	_	_
overspend	Past	0	56	1C-6	8	1	_	_	_	_	_	_	_
overshoot	Past	0	56	2C-7	25	2	_	_	_	_	_	_	_
wring	Past	0	50	2C-1	19	2	3A-2	34	3	wrang	_	2	increase
inbreed	Past	0	48	2C-6	24	2	_	_	_	_	_	_	_
smite	Past	0	46	3A-1	33	3	_	_	_	_	_	_	_
rethink	Past	2	42	2C-9	27	2	_	_	_	_	_	_	_
beget	Past	0	41	2D-5	32	2	2C-2	20	2	begat	3	151	_
underthrow	Past	0	35	2B-3	14	2	_	_	_	_	_	_	_
abide	Past	0	33	2C-8	26	2	_	_	_	_	_	_	_
outshine	Past	0	33	2C-8	26	2	_	_	_	_	_	_	_
knit	Past	0	31	1A-1	1	1	_	_	_	_	_	_	_
unsay	Past	0	31	2C-6	24	2	_	_	_	_	_	_	_
bespeak	Past	1	25	2D-4	31	2	_	_	_	_	_	_	_
unwind	Past	0	23	2C-5	23	2	_	_	_	_	_	_	_
redraw	Past	0	23	2B-4	15	2	_	_	_	_	_	_	_
unbind	Past	0	22	2C-5	23	2	_	_	_	_	_	_	_
forsake	Past	0	22	2B-1	12	2	_	_	_	_	_	_	_
unstick	Past	2	19	2C-1	19	2	_	_	_	_	_	_	_
burn	Past	0	20	1C-1	3	1	_	_	_	_	_	_	_
cowrite	Past	0	19	3A-1	33	3	_	_	_	_	_	_	_
shoe	Past	0	18	2C-7	25	2	_	_	_	_	_	_	_
miscast	Past	0	14	1A-1	1	1	_	_	_	_	_	_	_
recut	Past	0	13	1A-1	1	1	_	_	_	_	_	_	_
unmake	Past	0	12	1C-5	7	1	_	_	_	_	_	_	_
smell	Past	0	12	1C-2	4	1	_	_	_	_	_	_	_
foreknow	Past	0	11	2B-3	14	2	_	_	_	_	_	_	_
typeset	Past	0	11	1A-1	1	1	_	_	_	_	_	_	_
inlay	Past	0	11	1C-7	9	1	_	_	_	_	_	_	_
intercut	Past	0	11	1A-1	1	1	_	_	_	_	_	_	_
betake	Past	0	11	2B-1	12	2	_	_	_	_	_	_	_
heave	Past	0	10	2D-4	31	2	_	_	_	_	_	_	_
typecast	Past	0	9	1A-1	1	1	_	_	_	_	_	_	_
rerun	Past	0	9	2A-2	11	2	_	_	_	_	_	_	_
overfeed	Past	1	8	2C-6	24	2	_	_	_	_	_	_	_
uppercut	Past	0	9	1A-1	1	1	_	_	_	_	_	_	_
thrive	Past	0	7	3A-1	33	3	_	_	_	_	_	_	_
overwrite	Past	0	7	3A-1	33	3	_	_	_	_	_	_	_

wed	Past	0	7	1A-1	1	1	_	_	_	_	_	_	_
sting	Past	0	7	2C-1	19	2	_	_	_	_	_	_	_
chide	Past	0	6	2D-2	29	2	_	_	_	_	_	_	_
overblow	Past	0	5	2B-3	14	2	_	_	_	_	_	_	_
gird	Past	0	5	1C-6	8	1	_	_	_	_	_	_	_
overdraw	Past	0	4	2B-4	15	2	_	_	_	_	_	_	_
handwrite	Past	0	4	3A-1	33	3	_	_	_	_	_	_	_
unfreeze	Past	0	4	2D-4	31	2	_	_	_	_	_	_	_
bestride	Past	0	4	3A-1	33	3	_	_	_	_	_	_	_
bless	Past	0	3	1C-4	6	1	_	_	_	_	_	_	_
bereave	Past	0	3	2C-6	24	2	_	_	_	_	_	_	_
overhang	Past	0	3	2C-1	19	2	_	_	_	_	_	_	_
stave	Past	0	0	2C-8	26	2	_	_	_	_	_	_	_
strip	Past	0	0	1C-3	5	1	_	_	_	_	_	_	_
clap	Past	0	0	1C-3	5	1	_	_	_	_	_	_	_
overpay	Perfect	2	1461	1C-7	9	1	_	_	_	_	_	_	_
thrust	Perfect	5	886	1A-1	1	1	_	_	_	_	_	_	_
uphold	Perfect	2	528	2C-6	24	2	_	_	_	_	_	_	_
creep	Perfect	2	470	2C-6	24	2	_	_	_	_	_	_	_
shine	Perfect	4	256	2C-8	26	2	_	_	_	_	_	_	_
speed	Perfect	0	245	2C-6	24	2	_	_	_	_	_	_	_
rewrite	Perfect	5	765	3A-1	33	3	_	_	_	_	_	_	_
mistake	Perfect	30	3874	2B-1	12	2	_	_	_	_	_	_	_
overtake	Perfect	5	708	2B-1	12	2	_	_	_	_	_	_	_
pen	Perfect	0	0	1C-1	3	1	_	_	_	_	_	_	_
forecast	Perfect	1	569	1A-1	1	1	_	_	_	_	_	_	_
mislead	Perfect	8	500	2C-6	24	2	_	_	_	_	_	_	_
string	Perfect	3	670	2C-1	19	2	_	_	_	_	_	_	_
fling	Perfect	6	380	2C-1	19	2	_	_	_	_	_	_	_
undertake	Perfect	6	781	2B-1	12	2	_	_	_	_	_	_	_
cling	Perfect	0	122	2C-1	19	2	_	_	_	_	_	_	_
shrink	Perfect	4	805	3A-2	34	3	_	_	_	_	_	_	_
weave	Perfect	7	1445	2D-4	31	2	_	_	_	_	_	_	_
withhold	Perfect	6	636	2C-6	24	2	_	_	_	_	_	_	_
overthrow	Perfect	1	253	2B-3	14	2	_	_	_	_	_	_	_
stride	Perfect	0	0	3A-1	33	3	_	_	_	_	_	_	_
remake	Perfect	5	528	1C-5	7	2	_	_	_	_	_	_	_
plead	Perfect	0	123	2C-6	24	2	_	_	_	_	_	_	_
sweat	Perfect	0	0	1A-1	1	1	_	_	_	_	_	_	_
outgrow	Perfect	2	515	2B-3	14	2	_	_	_	_	_	_	_
kneel	Perfect	0	44	2C-6	24	2	_	_	_	_	_	_	_
withstand	Perfect	0	80	2C-3	21	2	_	_	_	_	_	_	_
inset	Perfect	0	23	1A-1	1	1	_	_	_	_	_	_	_
foresee	Perfect	1	335	2B-5	16	2	_	_	_	_	_	_	_
breed	Perfect	1	734	2C-6	24	2	_	_	_	_	_	_	_
stink	Perfect	1	258	3A-2	34	3	_	_	_	_	_	_	_
spell	Perfect	1	181	1C-2	4	2	_	_	_	_	_	_	_
forbid	Perfect	18	1960	2B-2	13	2	_	_	_	_	_	_	_
podcast	Perfect	0	12	1A-1	1	1	_	_	_	_	_	_	_
bend	Perfect	0	30	1C-6	8	1	_	_	_	_	_	_	_
offset	Perfect	5	979	1A-1	1	1	_	_	_	_	_	_	_
slit	Perfect	1	50	1A-1	1	1	_	_	_	_	_	_	_
recast	Perfect	0	261	1A-1	1	1	_	_	_	_	_	_	_
bust	Perfect	0	1	1A-1	1	1	_	_	_	_	_	_	_
babysit	Perfect	0	39	2C-2	20	2	_	_	_	_	_	_	_
befall	Perfect	1	101	2B-7	18	2	_	_	_	_	_	_	_
strive	Perfect	1	32	3A-1	33	3	_	_	_	_	_	_	_
rid	Perfect	2	476	1A-1	1	1	_	_	_	_	_	_	_

repay	Perfect	13	997	1C-7	9	1	_	_	_	_	_	_	_
sling	Perfect	2	173	2C-1	19	2	_	_	_	_	_	_	_
foretell	Perfect	1	100	2C-8	26	2	_	_	_	_	_	_	_
tread	Perfect	0	136	2D-5	32	2	_	_	_	_	_	_	_
outrun	Perfect	0	44	2A-2	11	2	_	_	_	_	_	_	_
dwell	Perfect	0	47	1C-2	4	1	_	_	_	_	_	_	_
wet	Perfect	0	0	1A-1	1	1	_	_	_	_	_	_	_
spoil	Perfect	0	375	1C-2	4	1	_	_	_	_	_	_	_
lean	Perfect	1	28	2C-6	24	2	_	_	_	_	_	_	_
behold	Perfect	0	28	2C-6	24	2	_	_	_	_	_	_	_
override	Perfect	1	110	3A-1	33	3	_	_	_	_	_	_	_
retell	Perfect	2	147	2C-8	26	2	_	_	_	_	_	_	_
outshoot	Perfect	0	59	2C-7	25	2	_	_	_	_	_	_	_
spill	Perfect	0	172	1C-2	4	1	_	_	_	_	_	_	_
overrun	Perfect	5	749	2A-2	11	2	_	_	_	_	_	_	_
bite	Perfect	5	1257	2D-2	29	2	_	_	_	_	_	_	_
partake	Perfect	2	76	2B-1	12	2	_	_	_	_	_	_	_
cleave	Perfect	0	1	2D-4	31	2	_	24	2	cleft	_	10	_
retake	Perfect	1	24	2B-1	12	2	_	_	_	_	_	_	_
oversleep	Perfect	0	20	2C-6	24	2	_	_	_	_	_	_	_
undercut	Perfect	0	111	1A-1	1	1	_	_	_	_	_	_	_
underwrite	Perfect	0	168	3A-1	33	3	_	_	_	_	_	_	_
ken	Perfect	0	111	1C-1	3	1	_	_	_	_	_	_	_
waylay	Perfect	0	106	1C-7	9	1	_	_	_	_	_	_	_
mishear	Perfect	1	51	2C-4	22	2	_	_	_	_	_	_	_
slink	Perfect	0	25	2C-1	19	2	_	_	_	_	_	_	_
rewind	Perfect	0	11	2C-5	23	2	_	_	_	_	_	_	_
overspend	Perfect	0	196	1C-6	8	1	_	_	_	_	_	_	_
overshoot	Perfect	0	36	2C-7	25	2	_	_	_	_	_	_	_
wring	Perfect	0	90	2C-1	19	2	_	_	_	_	_	_	_
inbreed	Perfect	0	83	2C-6	24	2	_	_	_	_	_	_	_
smite	Perfect	3	1016	3A-1	33	3	_	_	_	_	_	_	_
rethink	Perfect	0	68	2C-9	27	2	_	_	_	_	_	_	_
beget	Perfect	0	149	2D-5	32	2	_	_	_	_	_	_	_
underthrow	Perfect	0	4	2B-3	14	2	_	_	_	_	_	_	_
abide	Perfect	0	21	2C-8	26	2	_	_	_	_	_	_	_
outshine	Perfect	0	42	2C-8	26	2	_	_	_	_	_	_	_
knit	Perfect	10	3170	1A-1	1	1	_	_	_	_	_	_	_
unsay	Perfect	0	136	2C-6	24	2	_	_	_	_	_	_	_
bespeak	Perfect	0	1	2D-4	31	2	_	_	_	_	_	_	_
unwind	Perfect	0	84	2C-5	23	2	_	_	_	_	_	_	_
redraw	Perfect	0	0	2B-4	15	2	_	_	_	_	_	_	_
unbind	Perfect	0	30	2C-5	23	2	_	_	_	_	_	_	_
forsake	Perfect	2	378	2B-1	12	2	_	_	_	_	_	_	_
unstick	Perfect	0	78	2C-1	19	2	_	_	_	_	_	_	_
burn	Perfect	6	1196	1C-1	3	1	_	_	_	_	_	_	_
cowrite	Perfect	0	6	3A-1	33	3	_	_	_	_	_	_	_
shoe	Perfect	1	44	2C-7	25	2	_	_	_	_	_	_	_
miscast	Perfect	0	166	1A-1	1	1	_	_	_	_	_	_	_
recut	Perfect	0	10	1A-1	1	1	_	_	_	_	_	_	_
unmake	Perfect	0	31	1C-5	7	1	_	_	_	_	_	_	_
smell	Perfect	4	256	1C-2	4	1	_	_	_	_	_	_	_
foreknow	Perfect	0	5	2B-3	14	2	_	_	_	_	_	_	_
typeset	Perfect	2	37	1A-1	1	1	_	_	_	_	_	_	_
inlay	Perfect	0	30	1C-7	9	1	_	_	_	_	_	_	_
intercut	Perfect	5	42	1A-1	1	1	_	_	_	_	_	_	_
betake	Perfect	0	3	2B-1	12	2	_	_	_	_	_	_	_
heave	Perfect	0	9	2D-4	31	2	_	_	_	_	_	_	_

typecast	Perfect	2	129	1A-1	1	1	_	_	_	_	_	_	_
rerun	Perfect	0	0	2A-2	11	2	_	_	_	_	_	_	_
overfeed	Perfect	0	31	2C-6	24	2	_	_	_	_	_	_	_
uppercut	Perfect	0	3	1A-1	1	1	_	_	_	_	_	_	_
thrive	Perfect	0	0	3A-1	33	3	_	_	_	_	_	_	_
overwrite	Perfect	0	121	3A-1	33	3	_	_	_	_	_	_	_
wed	Perfect	4	649	1A-1	1	1	_	_	_	_	_	_	_
sting	Perfect	4	744	2C-1	19	2	_	_	_	_	_	_	_
chide	Perfect	0	0	2D-2	29	2	_	_	_	_	_	_	_
overblow	Perfect	0	4	2B-3	14	2	_	_	_	_	_	_	_
gird	Perfect	0	7	1C-6	8	1	_	_	_	_	_	_	_
overdraw	Perfect	0	240	2B-4	15	2	_	_	_	_	_	_	_
handwrite	Perfect	0	0	3A-1	33	3	_	_	_	_	_	_	_
unfreeze	Perfect	0	46	2D-4	31	2	_	_	_	_	_	_	_
bestride	Perfect	0	0	3A-1	33	3	_	_	_	_	_	_	_
bless	Perfect	0	52	1C-4	6	1	_	_	_	_	_	_	_
bereave	Perfect	0	0	2C-6	24	2	_	_	_	_	_	_	_
overhang	Perfect	0	7	2C-1	19	2	_	_	_	_	_	_	_
stave	Perfect	0	0	2C-8	26	2	_	_	_	_	_	_	_
strip	Perfect	0	0	1C-3	5	1	_	_	_	_	_	_	_
clap	Perfect	0	0	1C-3	5	1	_	_	_	_	_	_	_