



**MÁSTER EN ESTUDIOS INGLESES Y COMUNICACIÓN  
MULTILINGÜE E INTERCULTURAL**

## **TRABAJO FIN DE MÁSTER**

**A DIACHRONIC CORPUS-BASED STUDY OF NON-  
STANDARD PAST PARTICIPLES IN AMERICAN  
ENGLISH**

**de**

**ANDREA EROR**

**Realizado bajo la dirección de:  
Director: Dr. Javier Calle Martín**

**Departamento de Filología Inglesa, Francesa y Alemana  
Facultad de Filosofía y Letras  
UNIVERSIDAD DE MÁLAGA – 2021**



**MÁSTER EN ESTUDIOS INGLESES Y COMUNICACIÓN  
MULTILINGÜE E INTERCULTURAL**

## **TRABAJO FIN DE MÁSTER**

**A DIACHRONIC CORPUS-BASED STUDY OF NON-  
STANDARD PAST PARTICIPLES IN AMERICAN  
ENGLISH**

**UN ESTUDIO DIACRÓNICO BASADO EN CORPUS DE  
PARTICIPAOS PASADOS NO ESTÁNDARES EN INGLÉS  
AMERICANO**

**de**

**ANDREA EROR**

|   |  |                   |
|---|--|-------------------|
|  | <b>CALLE<br/>MARTIN<br/>JAVIER -<br/>33395976E</b><br>Firmado digitalmente por<br>CALLE MARTIN<br>JAVIER -<br>33395976E<br>Fecha: 2021.09.13<br>23:23:40 +02'00' |                   |
| Fdo. Andrea Eror  | Fdo. director/a  | Fdo. Codirector/a |

Málaga, a 13, de septiembre de 2021.

**Departamento de Filología Inglesa, Francesa y Alemana  
Facultad de Filosofía y Letras  
UNIVERSIDAD DE MÁLAGA – 2021**

**TITLE:** A DIACHRONIC CORPUS-BASED STUDY OF NON-STANDARD PAST PARTICIPLES IN AMERICAN ENGLISH

**KEYWORDS:** American English, Corpus Linguistics, non-standard, occurrence, past participle, variety.

**SUMMARY:**

This work studies selected forms of non-standard past participles in American English. *Gotten, proven, dreamed, spelled* and *burned* are diachronically studied in the *Corpus of Historical American English* (COHA). It starts with an overview on English as a global language, and emphasizes the importance of the historical and economic events that strengthened the supremacy and dominance of the American English variety. Following a quantitative approach and Corpus Linguistics, the main objectives of this study are to trace the origin of these forms, to study their development and to discern whether there is a possible existence of a process of Americanization in these forms. To do so, the data obtained from the COHA, will be compared with other corpora of other varieties of English, that is, the *British National Corpus* (BNC) for the British English variety, and the *Hong Kong Corpus of Spoken English* (HKCSE) for the Hong Kong English variety.

---

**TÍTULO:** UN ESTUDIO DIACRÓNICO BASADO EN CORPUS DE PARTICIPIOS DE PASADO NO ESTÁNDARES EN INGLÉS AMERICANO

**PALABRAS CLAVE:** Inglés Americano, Lingüística de Corpus, no estándar, ocurrencia, participio de pasado, variedad.

**RESUMEN:**

Este trabajo estudia formas seleccionadas de participios de pasado no estándares en inglés americano. *Gotten, proven, dreamed, spelled* and *burned* se estudian diacrónicamente en el *Corpus of Historical American English* (COHA). Comienza con una descripción general del inglés como idioma global y enfatiza la importancia de los eventos históricos y económicos que fortalecieron la supremacía y el dominio de la variedad del inglés americano. Siguiendo un enfoque cuantitativo, los principales objetivos de este estudio son rastrear el origen de estas formas, estudiar su desarrollo y comprobar la posible existencia de un proceso de americanización. Para ello, los datos obtenidos del COHA, serán comparados con otros corpus que representan diferentes variedades de inglés, esto es, el *British National Corpus* (BNC) para la variedad de inglés británico y el *Hong Kong Corpus of Spoken English* (HKCSE) para la variedad del inglés de Hong Kong.

## PLAGIARISM STATEMENT<sup>1</sup>

- I certify that this assignment/report is my own work, based on my personal study and/or research and that I have acknowledged all material and sources used in its preparation, whether they be books, articles, reports, lecture notes, and any other kind of document, electronic or personal communication.
- I also certify that this assignment/report has not previously been submitted for assessment in any other unit, except where specific permission has been granted from all unit coordinators involved, or at any other time in this unit, and that I have not copied in part or whole or otherwise plagiarised the work of other students and/or persons.

Student: Andrea Eror

Date: 13<sup>th</sup> September 2021

Signature:

Andrea Eror

---

<sup>1</sup> Adapted from <<http://people.uleth.ca/~selibl/Biol3200/Plagiarism.pdf>>

## Acknowledgments

Undertaking this master's degree has been a true challenge. At first, it was an unexplored area to me. Today, I confess to being a big enthusiast of English Studies.

First of all, I must express my heartfelt gratitude to Dr. Javier Calle Martín for introducing me to the boundless field of Linguistics and for infecting me with his enthusiasm. Thank you for believing in me and for giving me the opportunity to tackle some features of American English, one of my biggest passions. This work would have never been possible without your guidance and supervision.

Second, I have to thank my family for always being a colossal support for me. Mama, thank you for bringing me up with love and for showing me that I can reach my dreams. I would not be here today without you. Baba, thank you for giving me more love and support than I could ever return to you. Deda, thank you for always believing in me. And Miloš, thank you for constantly challenging me and helping me become the courageous person that I am today. I owe it all to you. *Volim vas.*

Also, I am thankful to all of my friends who have helped me, directly or indirectly. The same goes for the teachers that have guided me through my life; I will always remember you.

## Table of contents: Sections

|   |    |
|---|----|
| <b>1. INTRODUCTION</b>  | 1  |
| <b>2. THEORETICAL FRAMEWORK</b>   | 9  |
| <b>2.1 CORPUS-BASED STUDIES IN LANGUAGE</b>   | 9  |
| <b>2.2 CORPUS LINGUISTICS</b>   | 11 |
| <b>2.3. AMERICAN ENGLISH AND AMERICANIZATION</b>  | 14 |
| <b>3. RESEARCH METHODOLOGY</b>  | 18 |
| <b>4. RESULTS</b>   | 22 |
| <b>4.1. STATISTICAL SURVEY OF NON-STANDARD</b>  | 22 |
| <b>4.1.1. GOTTEN</b>  | 22 |
| <b>4.1.2. PROVEN</b>  | 24 |
| <b>4.1.3. DREAMED</b>   | 25 |
| <b>4.1.4. SPELLED</b>   | 27 |
| <b>4.1.5. BURNED</b>  | 28 |
| <b>4.2. STATISTICAL COMPARISON OF AMERICAN ENGLISH AND<br/>OTHER VARIETIES</b>          | 29 |
| <b>2.2.1. AMERICAN ENGLISH</b>  | 29 |
| <b>4.2.2. BRITISH ENGLISH</b>   | 30 |
| <b>4.2.3. HONG KONG ENGLISH</b>   | 31 |
| <b>5. DISCUSSION OF THE RESULTS</b>   | 33 |
| <b>5.1. AMERICAN ENGLISH</b>  | 33 |
| <b>5.2. THE CASES OF -ED SUFFIXES: <i>DREAMED</i>, <i>SPELLED</i> AND <i>BURNED</i></b> | 41 |
| <b>5.3. THE PROCESS OF AMERICANIZATION</b>  | 42 |
| <b>6. CONCLUSIONS</b>   | 45 |
| <b>REFERENCES</b>   | 48 |

## Table of contents: Tables

|   |    |
|---|----|
| 1. Table 1: <i>Got</i> and <i>gotten</i> in the COHA across 190 years     | 22 |
| 2. Table 2: <i>Proved</i> and <i>proven</i> in the COHA across 190 years  | 24 |
| 3. Table 3: <i>Dreamt</i> and <i>dreamed</i> in the COHA across 190 years | 26 |
| 4. Table 4: <i>Spelt</i> and <i>Spelled</i> in the COHA across 190 years  | 27 |
| 5. Table 5: <i>Burnt</i> and <i>burned</i> in the COHA across 190 years   | 28 |
| 6. Table 6: Standard and non-standard forms in the COHA                   | 30 |
| 7. Table 7: Standard and non-standard forms in the BNC                    | 30 |
| 8. Table 8: Standard and non-standard forms in the HKCSE                  | 31 |
| 9. Table 9: Normalised frequencies in all three corpora                   | 32 |

## Table of contents: Figures

|  |    |
|--|----|
| 1. Graph 1: <i>Got</i> and <i>gotten</i> in the COHA     | 34 |
| 2. Graph 2: <i>Proved</i> and <i>proven</i> in the COHA  | 35 |
| 3. Graph 3: <i>Dreamt</i> and <i>dreamed</i> in the COHA | 36 |
| 4. Graph 4: <i>Spelt</i> and <i>spelled</i> in the COHA  | 37 |
| 5. Graph 5: <i>Burnt</i> and <i>burned</i> in the COHA   | 39 |
| 6. Graph 6: Normalised data from the corpora             | 42 |



## 1. Introduction

English is an international language. It is widely known that English is considered to be the most popular language in the world. It is considered to be the one in which people can communicate everywhere. Nevertheless, it is not the most spoken language by people all around the world as a mother tongue and it is also not the official language in the biggest number of countries. So, why is it still considered a global language? David Crystal (2003) answers these questions in his book *English as a global language*, where he states that: “A language achieves a genuinely global status when it develops a special role that is recognized in every country” (p. 3). But what does Crystal refer to as a ‘special role’? He explains that in order for a language to be considered a global language, many countries must give it its special place in their communities. This means that English, from a long time ago, has been given a status of superiority. It has been taught in many countries as a foreign language and with time its knowledge became almost essential in many job positions due to tourism. So, in order for a language to be considered a global language, it is not mandatory for it to be the most spoken language in the world as a mother tongue nor to be the one that is spoken as the official language in many countries, as it happens with English.

Definitely, this language earned its special place a long time ago in countries all over the world and there are some historical events that have clearly influenced its international impact, from the British Empire and its colonies all around the world to the creation of a brand-new big country in 1776 whose official language happened to be English: the United States of America. Crystal (2003) argues that until the nineteenth century it was Britain who spread the English language at an international level due to

politics, but in the twentieth century the role was passed to a new economic leader: “During the twentieth century, this world presence was maintained and promoted almost single-handedly through the economic supremacy of the new American superpower. Economics replaced politics as the chief driving force” (p. 10). It is with the help of its industry, commerce, and agriculture that the U.S. gained its position as one of the new leading countries in the world: it is currently the world’s largest importer and the second largest exporter of goods. Also, it holds almost 30% of the world’s wealth. As a result of the increasing power of this country, even the diffusion of its official language grew. This has proved to be a clearly positive action for it, but there is still little controversy regarding the fact that it influences other varieties of English as well as the very structure of other languages. Different languages all over the world are taking borrowed words from English and there are also words that have been influenced or changed by it. There are many examples of this phenomenon, but the following examples are recognized and used internationally: *parking*, *weekend*, *camping*, *smoking*, and *jeans*. As a matter of fact, some of these popular borrowed words are funnily verbs in *-ing*, that are used in other languages as nouns. Furthermore, it must be mentioned that not only English has influenced other languages but other languages have influenced the English language. Many words from the ancient languages are in current use in English, such as *anonymous*, which is comes from the Greek *anōnumos*, and is defined as something or someone that has no name; or *avatar*, which comes from Sanskrit and is defined as the representation of someone in the virtual world, whose use also happened to increase with the release of the famous 2009 film, which has the same name. This is just one example of the influence of the cinematographic industry in language at an international level; this matter will indeed be discussed again further in. Nevertheless, it is important to keep in mind that these borrowed words in English not only come from ancient languages, but also from

present-day languages that have made a big influence in different areas overseas. For example, the French fashion industry has a big impact in the world of high fashion so that words as *catalogue* or *couture*, are often adapted to many other languages. So, it must be kept in mind that this phenomenon happens not only with English but with many other languages that have earned a big status or place in world's history or in a specific area, as it has been mentioned. Therefore, the globalization of the English language is just another proof of its supremacy.

This matter finally leads to the supremacy of American English over the past one hundred years or so, and the development of new technologies has done nothing but to increase its international influence. Given the fact that nowadays 80% of the world's electronic information is stored in English, and that most of the information online is also in this language, its expansion is practically unavoidable. One of the biggest worldwide impacts of American English that currently occurs is undeniably through the media. Its influence is at the highest of levels. It first began with economics, as Crystal (2003) states, but now a whole new era has developed and it is increasing each day. The U.S. is the official leader in the cinematographic, broadcasting, and music industry. Their music, films, and TV series are doubtlessly leading. In her study of slangs in both American English and British English, Hernando (2009) explains the importance of the cinematographic American production: "There is an increasing number of language learners depending on audiovisual material, which is a guaranteed source of speech input, even if the language that we find in cinematographic and television productions is not authentic discourse" (p. 28). Audiovisual material has in fact proven to be essential for language learners since they get to experience the real sounds and even to learn slang of a language through a screen. In fact, listening to the way a language sounds, is a great way of learning. The vocabulary that appears on the screen is spread at unimaginable

levels and its influence can be seen in the language. Not only English learners adopt it but also English speakers worldwide. That is why, the influence of the American English variety is increasing more and more. Overall, the influence of American English has increased from the twentieth century and nowadays its diffusion only continues to grow. Crystal (2003) summarizes it as follows:

We have already noted that the country contains nearly four times as many mother-tongue speakers of English as any other nation. It has been more involved with international developments in twentieth-century technology than any other nation. It is in control of the new industrial (that is, electronic) revolution. And it exercises a greater influence on the way English is developing worldwide than does any other regional variety. (p. 127).

The American English variety, with its own grammatical and lexical features, is showing its domination in the English language. It is convenient to say that as a result of this supremacy, its own words, verbs, phrases, slang, and many different features, American English is an attractive field for researchers all around the world. Therefore, the object of study of this work are non-standard past participles in American English over time (in the period 1820-2010) relying on the evidence found in the *Corpus of Historical English* (COHA). It must be mentioned that corpus-based studies in language have proved to be extremely useful since the corpus is a resourceful and useful tool when investigating many language features. A corpus (plural *corpora*) is a collection of texts, spoken or written, used as a language resource. Hernando (2009) defines its use as: “The use of corpora as a tool to carry out significant analysis of stretches of language, or alternatively (or simultaneously) of larger texts, can be both simple and resourceful” (p.28). Corpus-based studies allow the researcher to dig into features that interest him/her in language. Big corpora are often large amounts of data from different genres that are

usually specified, so that the results may be refined in terms of genre, register or other any likely speaker variation (age, gender, etc.). Viana, Zyngier and Barnbrook (2011) explained in their book that:

For Corpus Linguistics, language is in the first place something which is used rather than known, and the primary concern is to identify patterns of use in selected bodies of text from some population of language use, and the principles by which those patterns are constructed. (p. 3).

Corpus Linguistics is considered by many linguists to be a methodology as a method to obtain the needed results for a specific research, such as Davies. Other scholars, in turn, consider it to be a science or both a methodology and a science, such as Ashton. As it is known, language is in constant variation and this action cannot be disrupted, and it is possible to study language from its origin but it proves to be very difficult to make assumptions of its future. Nevertheless, Corpus Linguistics can help. Thanks to data obtained from different corpora it is definitely possible to observe tendencies in language and to determine possible future developments or tendencies for a specific phenomenon. This matter leads once again to the object of study of this work. As it is known, there is a big amount of interesting features regarding verbs in English that have been the object of study of many scholars such as Cheshire through the years. Moreover, when it comes to studying the non-standard forms of verbs in English there are many aspects that may seem attractive. First of all, there is a dilemma that must be addressed: it is believed that people commonly use the non-standard forms since they are unsure of the standard form that they turn to the one that sounds more natural. Cheshire (1993) addresses this issue in her article on standardization and irregular verbs in the following terms: “However, usage continues to be divided in the nonstandard varieties, and the majority of present-day native speakers of English still cannot be said to speak a regulated language” (p. 115). It

is true that both native and non-native speakers of English struggle with their use of this language. This is also caused by the large number of varieties of English that exist, such as British English, American English, Australian English, New Zealand English, and Hong Kong English, among others, which make things a little complicated for the speakers. Having contact with speakers of different varieties may cause someone to feel confused, especially in the case of non-native speakers. The survival of these non-standard forms is also troubled. But when they do, they invite investigation, just as Cheshire (1993) stated: “The variation that survives in standard English verbal morphology is so unusual that it invites investigation” (p.115). It is believed by many authors, such as Anderwald that the increasing popularity and use of the non-standard forms are due to their pronunciation or that it is due to ignorance of the ‘correct form’. For example, people find it easier to adopt the *-ed* ending for irregular verbs whose standard ending is *-t*, such as using *dreamed* instead of *dreamt* or *spelled* instead of *spelt*. It seems that for many speakers, it is more natural to use the non-standard form in everyday life than the standard one. This idea has been defended by Anderwald (2011) in her work, where she states that:

It has long been claimed that non-standard speech behaves in a more ‘natural’ way that the standardized varieties [...] but while this is well-documented for phonological features, it has been comparatively more difficult to demonstrate the same for non-standard features of morphology. (p. 251)

Overall, it has been proved to be of importance to study the development of non-standard past participles in American English in some selected irregular verbs, which in this case will be *gotten*, *proven*, *dreamed*, *spelled*, and *burned*. It is the aim of this work to study the reason why these forms have been increasingly popular throughout the years. To do so, as it has been mentioned, they will be diachronically studied through a period

of one hundred and ninety years, from 1820 to 2010, with the help of the *Corpus of Historical American English* (COHA). The main objectives of this study are: (a) to trace the origin of these forms in American English; (b) to analyze their development; and (c) to compare them with other varieties of English (also with the help of different corpora, such as the *British National Corpus* and the *Hong Kong Corpus of Spoken English*) in order to conclude if this phenomenon has been influenced by a process of Americanization or it may have happened randomly. The impact of the American industry and the diffusion of American English may have been of important influence regarding the development of these forms. Americanization is “the process of becoming more like the US or more like the people or the culture of the US”, according to the Cambridge Dictionary. But, is that what truly happened? This is a question that aims to be answered by the end of this meticulous research. Therefore, this work will mainly review some unique grammatical features in American English and their development, which will optimistically lead to relevant conclusions regarding the process of Americanization.

In addition to this brief introduction, it is convenient to mention the personal motivations that led to undertaking this research. Being born in the late twentieth century and growing up in the early 2000s, made one exposed to different kinds of content. With the accessibility of technology and the use of social media from a young age, also came the curiosity of learning foreign languages. As the years went by, time came in which languages have been chosen as an academic career. Studying a degree in translation and interpreting did nothing but increase the will of learning more about the way in which language works. The passion for the English language and being exposed to different kinds of American audiovisual material grew a special personal interest in American English. The individual research and interest about the unique features of this variety, started a long time ago, and it was therefore, the very first step that led to undertake this

research. Being currently a teacher of English, among other languages, made me want to know more about the way languages works. Moreover, the issue that has been addressed earlier about speakers not knowing which form of a specific word to use, happens to be a big issue and a frequently asked question when teaching English as a foreign language. Corpus Linguistics made possible a personal research in many different areas for me, and the origination of this dissertation itself. Also, being able to contribute in this area of research clearly feels like a personal and academic accomplishment. And, after this work, it is my intention to continue research in the area or in something related to it. Curiosity moves the world.



## 2. Theoretical framework

This chapter provides the theoretical framework of this work. First of all, it contains an overview of corpus-based studies in language. Afterwards, Corpus Linguistics is defined and described, and finally it addresses the topics of American English and Americanization. This chapter examines previous research that has been undertaken in the field of non-standard past participles in order to offer an insight in it. In addition, the research questions of this dissertation will be justified while light will be cast on these key concepts.

### 2.1. Corpus-based studies in Language

The use of corpora when studying language has been increasingly popular in the 21<sup>st</sup> century. Nevertheless, it is not a specifically new tool to study language. It is believed that the origins of corpora go back to the 13<sup>th</sup> century, with the Vulgate Bible. Even though their origins go way back and the different purposes that they have been used for, it is true that corpora are now a recurrent tool for more and more researchers. David Crystal (1995) gave a definition of corpora in his work *The Cambridge Encyclopedia of the English language*, where he states that:

Corpora are large and systematic enterprises: whole texts or whole sections of text are included, such as conversations, magazine articles, brochures, newspapers, lectures, sermons, broadcasts, and chapters of novels. Considerable thought is given to the selection of material so that, in the most general case, the corpus can stand as a reasonable representative sample of the language as a whole (a ‘general’ or ‘standard’ corpus). (p. 438).

In other words, corpora are considered a significantly good tool or methodology when researching in language. Through them, the researcher is able to detect different

aspects of the specific phenomenon that is being studied. Usually, it can help studying the evolution of language or studying language at a specific moment in time. Moreover, it can be used to study written language and spoken language, depending on the corpus. And, since the written material that has been selected to be part of the corpus is abundant and mixed (from different types of genres), the data obtained from it are considered to be representative for, for example, a type of variety. Cox (2011) agrees with it: “Within the context of corpus construction, decisions concerning the overall composition of a corpus and the eventual selection of its contents are often the result of careful corpus planning” (p. 245). This once again proves that the material that has been selected to be part of a corpus has not been haphazardly chosen. The researcher can rely on the material that he is working on as well as on the accuracy of the results. Even though the researcher must always be aware that the searching tools may not always be 100% accurate, it is always wise to check manually, if possible, that the correct results are given by the corpus when making a query. Corpora may also be used to compare and study different varieties of a language. There are different corpora that represent varieties of English, such as the ones that are the object of this study. For example, for historical and present-day American English, the *Corpus of Historical American English* (COHA), and for British English, the *British National Corpus* (BNC). Crystal (1995) also states that:

It enables investigation to make more objective and confident descriptions of usage than would be possible through introspection. It allows them to make statements about frequency of usage in the language as a whole, as well as comparative statements about usage in different varieties. It permits them, in principle, to arrive at a total account of the linguistic features in any of the texts contained in the corpus. And it provides them with a source of hypotheses about the way the language works. (p. 438).

Definitely, corpus-based studies are considered to be one of the best options when studying language. Because of the different tools that it offers when making a query, the researcher can view a lot of information on the occurrence of the word or phrase under scrutiny. For example, the corpus usually offers a KWIC view option, which allows to obtain the entire text where this specific word appears. Moreover, it provides information about its genre, year, and author. And, more importantly, it offers data of frequencies through the years and genres. Crystal thinks that these features that corpora allow for researchers to make ‘real’ statements and hypotheses on language use.

## 2.2. Corpus Linguistics

The previous argument leads to the concept of Corpus Linguistics. As it is known, it is the study of language as it is expressed in a corpus (a compilation of texts). That is why it has been given emphasis earlier to the importance of creating a well-planned and well-constructed corpora. The results of a specific research will mainly depend on the accomplishment of the corpora that is where the data are retrieved from. It is relevant to mention the contribution of *Perspectives on Corpus Linguistics* by V. Viana et al.(2011) in the field, where linguists and professionals have been interviewed, and each one of them has been asked the same questions regarding the origins of Corpus Linguistics, its use and its future.

In the very first chapter of this book, Guy Ashton, a professor of English Language and Translation from the University of Bologna, addresses different key concepts concerning Corpus Linguistics. As it has been said, its origins go back to the 13<sup>th</sup> century; Ashton (2011) states that “The first concordance was probably that compiled for the Vulgate Bible in the 13<sup>th</sup> century by Hugh of St. Cher, who employed 500 monks to list almost every word in the Bible with points where it was used” (p. 1). Since the very

beginning, the elaboration of corpora has proven to be a beneficial activity. As time went by, the development of technology has clearly had a big impact on every area of research, and Corpus Linguistics are not an exception. What earlier implied extremely meticulous and hard work, today can be done within seconds. And that is probably the biggest advantage of corpora: they are fast. Big well-known corpora, such as the COHA, have new updates almost every year. They are constantly fixing errors as well as improving the speed. A query can be made within seconds and they are very easy to use. This fact clearly helps and attracts researchers to dive into corpus-based studies.

But is Corpus Linguistics considered a science or a methodology? Both views are hold in this dissertation. Even though this is a common dilemma and the aim here is not to impose one theory or another, it is going to be explained. In this particular work, Corpus Linguistics is used as a methodology to answer specific questions concerning the non-standard form of past participles. But Corpus Linguistics has also proven to be a science and the object of study for many researches around the world. This issue was also addressed by Ashton (2011) in his interview, who elaborated that:

Given its predominant concern with applications, Corpus Linguistics must be viewed as a methodology, whether employed to provide data for dictionaries and grammars, to produce language teaching syllabuses and materials, or for natural language processing procedures of speech recognition, automatic text categorization/summarization, machine translation, or authorship attribution. On the other hand, Corpus Linguistics is a science inasmuch as it has a particular object of study, namely language as it is actually used in naturally-occurring speech and writing. (p. 2).

Ashton explains the different purposes that corpora have as a methodology, and agrees that it must also be considered as a science, especially due to its recent popularity

in language research. It is known that Corpus Linguistics is not only important for the linguistic community but also for the research community. Apart from these specific purposes, this discipline is also frequently used to compare varieties of a language. Pereira Da Costa (2019) explains the importance of parallel corpora in his doctoral thesis on multilingual data collection for different corpus-based approaches to translation:

High-quality parallel corpora are a preferred resource in the language engineering and the linguistics communities. Nevertheless, the lack of sufficient and up-to-date parallel corpora, especially for narrow domains and poorly-resourced languages is currently one of the major obstacles to further advancement across various areas like translation, language learning and automatic and assisted translation. (p. 33).

He mentions the importance of parallel corpora in research, since this is a commonly used tool to study different varieties, as it has just been mentioned. It is true that if more parallel corpora developed, it would perhaps seem even more attractive for scholars to undertake research in this field since it would clearly make their purpose easier. As a way of concluding this brief introduction to Corpus Linguistics the issue of its future must not stay unnoticed. What will happen in the future with the use of corpora is a common question. More specifically, there are many concerns about what will happen in the near future of publicly-available corpora. Going back to V. Viana et al.'s (2011) book, Mark Davies, the creator of the *Corpus of Historical American English* (which is the object of this study) and many others, have been interviewed. When he is asked to express his opinion on future corpora he stated that: "The two fundamental problems with it are (a) who is going to fund the corpus, and (b) who is going to provide the texts" (p. 75). The corpora that will be meticulously studied in this work are going to be explained in Chapter 3. But they all have an important thing in common: they have all been created with the help of grants. The importance of investing on the future of research and science is

essential. And for the creation of future corpora the funding is truly necessary. With appropriate funding and the work of professionals, there is hope in the future development of this field.

### **2.3. American English and Americanization**

American English is the object of study of this work. A brief introduction to this variety has been provided in the introduction (Chapter 1). It is important to understand some key concepts in order to study the development and present-day status of American English. Crystal talks about the ‘creation’ of the American English variety. He brings attention to Noah Webster and his contribution in the following terms (1995):

In his work, Webster proposed the institution of an ‘American standard’. It was partly a matter of honour as an independent nation... to have a system of our own, in language as well as government’; it was partly a matter of common sense, because in England ‘the taste of her writers is already corrupted, and her language on the decline; and it was partly a matter of practicality, England being at ‘too great distance to be our model’. (p. 80).

Webster was committed to give his country a level of supremacy and a level of distinction, so he thought it was necessary to ‘create’ a language of their own: American English. In his attempt to do so, in 1828 he compiled *An American Dictionary of the English Language*. Crystal (1995) explains that: “The work greatly improved the coverage of scientific and technical terms, as well as terms to do with American culture and institutions (such as congress and plantation), and added a great deal of encyclopedic information” (p .81). Even though it was a positive thing for American English to be represented in such a way, later Webster came to the realization that there were not more than fifty words which were added that were different to British English. As time passed

by, the vocabulary of American English developed and increased but it is true that during the 19<sup>th</sup> century there was also a dilemma concerning the American literary production and the one that was taking place in Europe. Even though many texts were produced, just a few obtained a high number of sales. Crystal explains that it was in these works that “would appear the results of the vast tide of lexical innovation which was already, in those early decades, transforming the linguistic identity of the new nation” (p. 83). During this new era, the American vocabulary was influenced not only by Spanish and native American words, but also with old English words that adopted new senses and meanings (Crystal, 1995, p. 83). Overall, this was the very beginning of what is nowadays known as American English. After centuries of development, this variety nowadays has proven to be of relevant influence on the English language. Since their economic supremacy in the 20<sup>th</sup> century and its predominance in technology and media in the 21<sup>st</sup> century, American English happens to be an object of study for many researchers: it is interesting to study how this variety developed from ‘standard’ English, to what it is today, with its big number of specific features.

*Americanism* is a term that was used for the first time by Witherspoon (1781), and it can be defined as a number of values that conform the American identity, which includes its traditions, politics, and culture, among others. With *Americanism*, also came *Americanization*. Which is, as mentioned earlier, the process of something becoming American. But is that what happened to the British English variety when the U.S. developed as a country? The answer is clearly yes. Gonçalves et al. (2018) explained it in their work *Mapping the Americanization of English in space and time*:

In fact, the “Americanization” of (global) English is one of the main processes of language change in contemporary English. Although it is found to work along with other processes such as colloquialization and informalization the spread of American

features all over the globe is generally assumed to be result of the American ‘leadership’ in change (p. 2).

The term *globalization* is also linked to this idea. Even though it is known more as an economic process, this term can also be applied to language. As it is known, English is considered to be a global language and the influence of the American variety due to its supremacy or leadership, has definitely contributed to it. That is why this variety has especially increased the interest of researchers. This leads to the object of this work, which is to study selected forms of non-standard past participles in American English over time with the help of a corpus (COHA). One of the main objectives is to determine whether a process of Americanization has been present in the development of these forms. In what concerns previous research on non-standard forms, there is one work of Wikström, from the University of Gothenburg, that must be highlighted. In this work, he studied the non-standard preterit and past participle *-ed* forms in present-day American English. It was done with the help of the *Corpus of Contemporary American English* (COCA), focusing on the verbs *burn*, *dream*, *learn*, *spell*, *spill* and *spoil*. Wikström (2013) states that: “Present-day AmE prefers the *-ed* suffix over the *-t* suffix for the past participle of this group of verbs” (p. 21). He found that the non-standard forms with the suffix *-ed* were mainly used in AmE, nevertheless he encourages further and more specific research in this area: “A contrastive corpus-based analysis of this varied use in two varieties of English, BrE and AmE, can be suggested as a topic of further research in order to see statistical tendencies in this type of variety- specific variation” (p. 21).

For this work, Crystal’s division in classes of irregular verbs depending on their features and tendencies of use, will be used:

- Class 1: formed by verbs whose irregular feature is their ending for both their past



and participle forms. For example, *send/sent*.

- Class 2: formed by the verbs whose *-ed* participle forms have both *-n* and *-ed* endings, such as *swollen/swelled*.
- Class 3: verbs that have the same ending for past and *-ed* participle forms such as, *keep/kept*.
- Class 4: verbs that have an *-n* ending for the *-ed* participle form and an irregular past form, such as *blow/blew/blown*.
- Class 5: verbs that always have the same form always, such as *cut*.
- Class 6: verbs that have no ending but that use the same form for past and participle, for example, *sit/sat*.
- Class 7: verbs that have no ending, that their past and participle forms are different, and whose vowels change, such as *go/went/gone*.

In this dissertation, verbs from Class 2 will be studied (*proven, dreamed, spelled* and *burned*), as well as a verb from Class 5 (*gotten*).

### 3. Research methodology

The diachronic corpus-based study of selected forms of non-standard past participles in American English has different parts that must be addressed. An empirical methodology based on a quantitative method will be used and the topic will be addressed both diachronically and diatopically:

- a) to trace the origin of these forms in American English
- b) to analyze their development<sup>[1]</sup><sub>[SEP]</sub>over time
- c) to compare them with other varieties of English (also with the help of different corpora, such as the BNC Web) in order to conclude if this phenomenon has been influenced by a process of Americanization

On methodological grounds, the corpus used as source of evidence for the study is the *Corpus of Historical American English* (COHA). Compiled by Mark Davies, it was finally released in 2010. It is the largest structured corpus of historical English with more than 475 million words of text from the 1820s to the 2010s. In view of its time span, this corpus is also suitable for researching changes in current language. When compared to other corpora of historical AmE, the COHA has proven to be a hundred times larger than any of the existing corpora. That is why it is considered to be the most complex one, as well as the chosen input by many scholars for the investigation of a particular linguistic feature. The data from COCA come from a variety of five genres:

- TV/Movies. This genre contains more than 40 million words from TV and movies subtitles and transcripts.
- Fiction. With more than 222 million words, this genre offers different data from books, movies, play scripts, and data from COCA.
- Magazine. A compilation of texts from magazines from 1810 to 2010 with more than 106 million words. Each decade has at least ten different magazines with different sub-genres.
- Newspaper. It contains more than 45 million words of text from historical archives

of newspapers from 1850 to 1980, as well as data from COCA concerning this genre which go from 1990 to 2010.

- Non-fiction. With more than 61 million words, this genre offers a compilation of different libraries and its books from 1810 to 2010.

The corpus is balanced in terms of its genres and subgenres and this fact assures truthful results for researchers. The COHA contains 475 million words of text (475,031,831 words to be precise) which will all be processed for the present study. In the light of this piece of evidence, the COHA is considered to be the most useful corpus when studying the development of historical AmE as well as ongoing changes in current AmE. In this particular case, we are concerned with the origin and development of the selected non-standard past participles in American English. The competition of standard and non-standard past participles will also be investigated, in our case, *got* and *gotten*, *proved* and *proven*, *dreamt* and *dreamed*, *spelt* and *spelled*, and *burnt* and *burned*. For the sake of normalisation, It is important to keep in mind the different length of the decades that will be studied, for the sake of comparison. There are the following:

- 1820-1830: 6,981,389 words
- 1830-1840: 13,711,287 words
- 1840-1850: 15, 807, 047 words
- 1850-1860: 16,536,003 words
- 1860-1870: 16,936,003 words
- 1870-1880: 18,788,467 words
- 1880-1890: 20,067,205 words
- 1890-1900: 20,426,783 words
- 1900-1910: 21,977,250 words
- 1910-1920: 23,103,098 words
- 1920-1930: 25,700,422 words
- 1930-1940: 27,707,879 words
- 1940-1950: 27,399,750 words
- 1950-1960: 28,661,274 words

- 1960-1970: 29,122,676 words
- 1970-1980: 28,829,225 words
- 1980-1990: 29,851,580 words
- 1990-2000: 33,149,318 words
- 2000-2010: 34,821,812 words
- 2010-2020: 35,452,806 words

The selected standard and non-standard past participles, *got* and *gotten*, *proved* and *proven*, *dreamt* and *dreamed*, *spelt* and *spelled*, and *burnt* and *burned*, are going to be studied from the beginning till the end of the period of time in COHA, which is from 1820 to 2010. It has been chosen to study this aspect during a long period of time in order to obtain enough data to fulfill the objectives of this study, which are to trace their origin, development, and discern any likely type of variation.

Apart from the COHA, the current use of these verb forms have been surveyed in other corpora. First, the data for the use and distribution of these forms in current American British English come from the *Corpus of Contemporary American English* (COCA) and the *British National Corpus* (BNC), respectively. On the one hand COCA, also created by Mark Davies, contains a total number of 1,002,889,754 words each year from 1990 to 2019. This corpus has been chosen since it is representative of present-day AmE. It contains texts from eight different genres: (a) spoken; (b) fiction; (c) magazines; (d) newspapers; (e) academic; (f) web (genl); (g) web (blog); and (h) TV/movies. On the other hand, BNC was originally created in the early 1990s by Oxford University press. Its newest edition was released in 2007, named *BNC XML Edition*. The BNC is considered to be representative of British English since it provides data from a wide range of sources. Its written part represents the 90% of the corpus, while its spoken part consists of a 10% of it. It contains a total of 96,263,399 words in seven different genres: (a) spoken, which contains transcripts of conversations; (b) fiction, containing drama, poetry,

and prose; (c) magazine; (d) newspaper, containing different newspaper with a wide number of subgenres; (e) non-academic; (f) academic, containing data from engineering, arts, law, medicine and sciences; and (g) miscellaneous.

Second, the diatopic analysis relies on the *Hong Kong Corpus of Spoken English* (HKCSE). It is hosted by the Research Centre for Professional Communication in English of the Hong Kong Polytechnic University. Like Mark Davies' corpora, the creation of HKCSE was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region (Project No. G-YE86). It currently contains 907,657 words divided into four genres: (a) academic sub-corpus, which contains lectures, seminars, tutorials, student presentation and Q&A, writing Assistance Programme consultation, and workshop for academic and research staff; (b) business sub-corpus, formed by job and placement interviews, presentations, meetings, informal office talk, announcements, Q&A's, presentations, conference talks, workplace telephone talk, and service encounters; (c) conversation sub-corpora, which contains conversations collected in a big range of social settings; (d) public sub-corpora, which contains speeches, interviews, press briefings, Q&A's, discussion forums, and radio announcements.

## 4. Results

This chapter presents the results of our analysis. This chapter is divided into two sections. First, a statistical survey of the selected forms of non-standard past participles in the COHA will be presented over time. In this section, the data concerning the standard and non-standard forms from this corpus will be represented. In the tables below, both absolute and normalised figures will be represented due to the different size of each time period. Second, a statistical comparison of AmE and two other varieties will be shown, that is, BrE and HKE.

### 4.1. Statistical survey of the selected forms of non-standard past participles in the COHA

In this section, the data retrieved from the *Corpus of Historical American English* (COHA) of the non-standard forms (*gotten*, *proven*, *dreamed*, *spelled* and *burned*) as well as their standard forms (*got*, *proved*, *dreamt*, *spelt* and *burnt*) will be presented in different tables. The absolute frequencies in each decade will be represented as well as their normalised frequencies. To obtain these, the following formula will be used: raw frequency \* common base (100,000 in this case) / total number of words of each decade. Each verb will be represented in its own sub-section below.

#### 4.1.1. Gotten

First, the distribution of *got* vs *gotten* is presented in Table 1 in terms of the different decades in which the corpus is organized.

|             | <b>GOT<br/>ABSOLUTE</b> | <b>GOTTEN<br/>ABSOLUTE</b> | <b>GOT<br/>NORMALISED</b> | <b>GOTTEN<br/>NORMALISED</b> |
|-------------|-------------------------|----------------------------|---------------------------|------------------------------|
| <b>1820</b> | 688                     | 29                         | 9,85                      | 0,41                         |

|              |        |       |         |       |
|--------------|--------|-------|---------|-------|
| <b>1830</b>  | 2956   | 34    | 21,55   | 0,24  |
| <b>1840</b>  | 3275   | 37    | 20,71   | 0,23  |
| <b>1850</b>  | 4562   | 38    | 27,58   | 0,22  |
| <b>1860</b>  | 5618   | 78    | 33,16   | 0,46  |
| <b>1870</b>  | 7265   | 100   | 38,66   | 0,53  |
| <b>1880</b>  | 8993   | 107   | 44,02   | 0,52  |
| <b>1890</b>  | 8080   | 151   | 39,55   | 0,73  |
| <b>1900</b>  | 11757  | 202   | 53,49   | 0,91  |
| <b>1910</b>  | 16735  | 253   | 72,43   | 1,09  |
| <b>1920</b>  | 17039  | 315   | 66,29   | 1,22  |
| <b>1930</b>  | 32726  | 372   | 118,11  | 1,34  |
| <b>1940</b>  | 34442  | 458   | 125,70  | 1,67  |
| <b>1950</b>  | 36401  | 718   | 127     | 2,5   |
| <b>1960</b>  | 35054  | 955   | 120,36  | 3,27  |
| <b>1970</b>  | 40445  | 1059  | 140,29  | 3,67  |
| <b>1980</b>  | 41304  | 1306  | 138,36  | 4,37  |
| <b>1990</b>  | 41779  | 1964  | 126,03  | 5,92  |
| <b>2000</b>  | 41447  | 2429  | 119,02  | 6,97  |
| <b>2010</b>  | 38932  | 2678  | 109,81  | 7,55  |
| <b>TOTAL</b> | 429498 | 13283 | 1551,97 | 43,82 |

Table 1: *Got* and *gotten* in the COHA across 190 years

The verb *to get* is clearly one of the most frequent verbs in English. Therefore, it has not been chosen haphazardly for this study. The fact is that the non-standard form has been increasingly used over time, *got* as the standard form and *gotten*

as the non-standard one. As it can be observed in the table, the occurrences of both forms have increased over the decades. In 1820, the form *gotten* had only 0,41 occurrences in a corpus of 100,000 words, while in 2010 it had 7,55 occurrences. It can be seen that its use has notably increased. Nevertheless, the data show the preference for the standard form *got* over *gotten*. These, and all of the remaining forms, will be meticulously discussed in Chapter 5.

#### 4.1.2. Proven

Second, the data concerning the forms *proved* and *proven* are shown.

|             | <b>PROVED<br/>ABSOLUTE</b> | <b>PROVEN<br/>ABSOLUTE</b> | <b>PROVED<br/>NORMALISED</b> | <b>PROVEN<br/>NORMALISED</b> |
|-------------|----------------------------|----------------------------|------------------------------|------------------------------|
| <b>1820</b> | 664                        | 3                          | 9,51                         | 0,04                         |
| <b>1830</b> | 1185                       | 13                         | 8,64                         | 0,09                         |
| <b>1840</b> | 1347                       | 25                         | 8,52                         | 0,15                         |
| <b>1850</b> | 1518                       | 37                         | 9,17                         | 0,22                         |
| <b>1860</b> | 1559                       | 54                         | 9,2                          | 0,31                         |
| <b>1870</b> | 1613                       | 69                         | 8,58                         | 0,36                         |
| <b>1880</b> | 1606                       | 88                         | 8,003                        | 0,43                         |
| <b>1890</b> | 1608                       | 83                         | 7,87                         | 0,40                         |
| <b>1900</b> | 1783                       | 117                        | 8,11                         | 0,53                         |
| <b>1910</b> | 1734                       | 124                        | 7,71                         | 0,50                         |
| <b>1920</b> | 1726                       | 116                        | 6,71                         | 0,45                         |
| <b>1930</b> | 1421                       | 113                        | 5,12                         | 0,40                         |
| <b>1940</b> | 1603                       | 104                        | 5,85                         | 0,37                         |



|              |       |      |        |       |
|--------------|-------|------|--------|-------|
| <b>1950</b>  | 1733  | 135  | 6,04   | 0,47  |
| <b>1960</b>  | 1512  | 163  | 5,19   | 0,55  |
| <b>1970</b>  | 1232  | 248  | 4,27   | 0,86  |
| <b>1980</b>  | 1232  | 277  | 4,12   | 0,92  |
| <b>1990</b>  | 1050  | 356  | 3,16   | 1,07  |
| <b>2000</b>  | 1163  | 436  | 3,33   | 1,25  |
| <b>2010</b>  | 1054  | 519  | 2,97   | 1,46  |
| <b>TOTAL</b> | 28407 | 3080 | 132,07 | 10,83 |

Table 2: *Proved* and *proven* in the COHA across 190 years

At first sight, it can be seen that both forms *proved* and *proven* have increased over time. Nevertheless, the standard form *proved* clearly predominates in this case with a total number of 132,07 occurrences, while the non-standard form *proven* has a total number of 10,83 occurrences, in a corpus of 100,000 words. The standard form in this case predominates, as it occurred in the previous subsection with the forms *got* and *gotten*.

#### 4.1.3. Dreamed

Third, the data concerning *dreamt* and *dreamed* is going to be shown. In the following table the occurrences retrieved from the corpus are represented:

|             | <b>DREAMT<br/>ABSOLUTE</b> | <b>DREAMED<br/>ABSOLUTE</b> | <b>DREAMT<br/>NORMALISED</b> | <b>DREAMED<br/>NORMALISED</b> |
|-------------|----------------------------|-----------------------------|------------------------------|-------------------------------|
| <b>1820</b> | 45                         | 69                          | 0,64                         | 0,98                          |
| <b>1830</b> | 34                         | 233                         | 0,24                         | 1,69                          |
| <b>1840</b> | 30                         | 317                         | 0,18                         | 2                             |

|              |      |       |      |       |
|--------------|------|-------|------|-------|
| <b>1850</b>  | 20   | 432   | 0,12 | 2,61  |
| <b>1860</b>  | 28   | 462   | 0,16 | 2,72  |
| <b>1870</b>  | 49   | 524   | 0,26 | 2,78  |
| <b>1880</b>  | 44   | 596   | 0,21 | 2,97  |
| <b>1890</b>  | 42   | 595   | 0,2  | 2,91  |
| <b>1900</b>  | 35   | 646   | 0,15 | 2,93  |
| <b>1910</b>  | 38   | 589   | 0,16 | 2,54  |
| <b>1920</b>  | 36   | 681   | 0,14 | 2,64  |
| <b>1930</b>  | 50   | 563   | 0,18 | 2,03  |
| <b>1940</b>  | 46   | 535   | 0,16 | 1,95  |
| <b>1950</b>  | 37   | 553   | 0,12 | 1,92  |
| <b>1960</b>  | 47   | 531   | 0,16 | 1,82  |
| <b>1970</b>  | 95   | 496   | 0,32 | 1,72  |
| <b>1980</b>  | 69   | 564   | 0,23 | 1,88  |
| <b>1990</b>  | 104  | 682   | 0,31 | 2,05  |
| <b>2000</b>  | 89   | 662   | 0,25 | 1,90  |
| <b>2010</b>  | 97   | 545   | 0,27 | 1,53  |
| <b>TOTAL</b> | 1035 | 10275 | 4,46 | 43,57 |

Table 3: *Dreamt* and *dreamed* in the COHA across 190 years

It must be mentioned the results are significantly different. The standard form *dreamt* amounts to 4,46 occurrences, being much smaller than the total number of occurrences of the non-standard form *dreamed*, with a total of 43,57 occurrences. Even though the use of both forms varies over time, there is a clear preference for the non-standard form in this case. The reason why this happens will be discussed in Chapter 5.

#### 4.1.4. Spelled

The forms *spelt* and *spelled* are shown below.

|             | <b>SPELT<br/>ABSOLUTE</b> | <b>SPELLED<br/>ABSOLUTE</b> | <b>SPELT<br/>NORMALISED</b> | <b>SPELLED<br/>NORMALISED</b> |
|-------------|---------------------------|-----------------------------|-----------------------------|-------------------------------|
| <b>1820</b> | 5                         | 1                           | 0,07                        | 0,01                          |
| <b>1830</b> | 8                         | 11                          | 0,05                        | 0,08                          |
| <b>1840</b> | 12                        | 15                          | 0,07                        | 0,09                          |
| <b>1850</b> | 16                        | 26                          | 0,09                        | 0,15                          |
| <b>1860</b> | 26                        | 33                          | 0,15                        | 0,19                          |
| <b>1870</b> | 11                        | 55                          | 0,05                        | 0,29                          |
| <b>1880</b> | 18                        | 58                          | 0,08                        | 0,28                          |
| <b>1890</b> | 14                        | 61                          | 0,06                        | 0,29                          |
| <b>1900</b> | 10                        | 94                          | 0,04                        | 0,42                          |
| <b>1910</b> | 29                        | 102                         | 0,12                        | 0,44                          |
| <b>1920</b> | 14                        | 109                         | 0,05                        | 0,42                          |
| <b>1930</b> | 9                         | 117                         | 0,03                        | 0,42                          |
| <b>1940</b> | 11                        | 133                         | 0,04                        | 0,48                          |
| <b>1950</b> | 8                         | 227                         | 0,02                        | 0,79                          |
| <b>1960</b> | 10                        | 170                         | 0,03                        | 0,58                          |
| <b>1970</b> | 15                        | 184                         | 0,05                        | 0,63                          |
| <b>1980</b> | 1                         | 151                         | 0,003                       | 0,50                          |
| <b>1990</b> | 10                        | 169                         | 0,03                        | 0,50                          |
| <b>2000</b> | 6                         | 174                         | 0,01                        | 0,49                          |

|              |     |      |      |      |
|--------------|-----|------|------|------|
| <b>2010</b>  | 7   | 167  | 0,01 | 0,47 |
| <b>TOTAL</b> | 240 | 2057 | 0,98 | 7,51 |

Table 4: *Spelt* and *spelled* in the COHA across 190 years

At first glance, as it happened with the forms *dreamt* and *dreamed*, the results are also significantly different. The total number of occurrences of what is considered to be the standard form *spelt* is 0,98 while the total number of occurrences of the non-standard form *spelled* is 7,51 occurrences. The use of both forms varies over time but the non-standard form is clearly more frequent in the corpus.

#### 4.1.5. Burned

Finally, the data concerning the forms *burnt* and *burned* are shown.

|             | <b>BURNT<br/>ABSOLUTE</b> | <b>BURNED<br/>ABSOLUTE</b> | <b>BURNT<br/>NORMALISED</b> | <b>BURNED<br/>NORMALISED</b> |
|-------------|---------------------------|----------------------------|-----------------------------|------------------------------|
| <b>1820</b> | 176                       | 86                         | 2,5                         | 1,23                         |
| <b>1830</b> | 188                       | 223                        | 1,37                        | 1,62                         |
| <b>1840</b> | 272                       | 417                        | 1,72                        | 2,63                         |
| <b>1850</b> | 249                       | 484                        | 1,50                        | 2,92                         |
| <b>1860</b> | 269                       | 646                        | 1,58                        | 3,81                         |
| <b>1870</b> | 275                       | 672                        | 1,46                        | 3,57                         |
| <b>1880</b> | 250                       | 790                        | 1,25                        | 3,93                         |
| <b>1890</b> | 187                       | 842                        | 0,91                        | 4,12                         |
| <b>1900</b> | 178                       | 954                        | 0,80                        | 4,34                         |
| <b>1910</b> | 173                       | 849                        | 0,74                        | 3,67                         |
| <b>1920</b> | 207                       | 1094                       | 0,80                        | 4,25                         |

|              |      |       |      |      |
|--------------|------|-------|------|------|
| <b>1930</b>  | 236  | 1187  | 0,85 | 4,28 |
| <b>1940</b>  | 194  | 1204  | 0,70 | 4,39 |
| <b>1950</b>  | 174  | 1188  | 0,60 | 4,14 |
| <b>1960</b>  | 167  | 1066  | 0,57 | 3,66 |
| <b>1970</b>  | 135  | 1047  | 0,46 | 3,63 |
| <b>1980</b>  | 192  | 1139  | 0,64 | 3,81 |
| <b>1990</b>  | 236  | 1305  | 0,79 | 3,93 |
| <b>2000</b>  | 237  | 1379  | 0,68 | 3,96 |
| <b>2010</b>  | 280  | 1245  | 0,78 | 3,51 |
| <b>TOTAL</b> | 4275 | 17817 | 20,7 | 71,4 |

Table 5: *Burnt* and *burned* in the COHA across 190 years

At first sight, as it happened with the issue of the previous two forms, the results of *burnt* and *burned* are also significantly different. The total number of occurrences of the standard form *burnt* is 20,7 while the total number of occurrences of the non-standard form *burned* is 71,4 occurrences. The non-standard *burned* is clearly much more frequent in the corpus.

## 4.2. Statistical comparison of American English and other varieties

The previous data are exclusively concerned with AmE. Now the comparison will be made with other corpora.

### 4.2.1. American English

Regarding the AmE variety, the *Corpus of Contemporary American English* (COCA) has been chosen. The raw frequencies concerning this corpus are represented in the

following table:

|                | <b>TOTAL<br/>OCCURRENCES</b> |
|----------------|------------------------------|
| <b>GOT</b>     | 1059368                      |
| <b>GOTTEN</b>  | 57904                        |
| <b>PROVED</b>  | 28341                        |
| <b>PROVEN</b>  | 21519                        |
| <b>DREAMT</b>  | 1903                         |
| <b>DREAMED</b> | 10255                        |
| <b>SPELT</b>   | 330                          |
| <b>SPELLED</b> | 4751                         |
| <b>BURNT</b>   | 5388                         |
| <b>BURNED</b>  | 24868                        |

Table 6: Standard and non-standard forms in the COCA

#### 4.2.2. British English

The data concerning the BrE variety is obtained from the *British National Corpus* (BNC). The raw frequencies data are represented in the following table:

|               | <b>TOTAL<br/>OCCURRENCES</b> |
|---------------|------------------------------|
| <b>GOT</b>    | 89430                        |
| <b>GOTTEN</b> | 103                          |
| <b>PROVED</b> | 6725                         |
| <b>PROVEN</b> | 780                          |

|                |      |
|----------------|------|
| <b>DREAMT</b>  | 268  |
| <b>DREAMED</b> | 716  |
| <b>SPELT</b>   | 436  |
| <b>SPELLED</b> | 252  |
| <b>BURNT</b>   | 1246 |
| <b>BURNED</b>  | 1408 |

Table 7: Standard and non-standard forms in the BNC

### 4.2.3. Hong Kong English

The data concerning the variety of HKE is retrieved from the *Hong Kong Corpus of Spoken English* (HKCSE). The raw frequencies are represented in the following table:

| TOTAL OCCURRENCES |      |
|-------------------|------|
| <b>GOT</b>        | 1186 |
| <b>GOTTEN</b>     | 5    |
| <b>PROVED</b>     | 5    |
| <b>PROVEN</b>     | 7    |
| <b>DREAMT</b>     | 0    |
| <b>DREAMED</b>    | 0    |
| <b>SPELT</b>      | 0    |
| <b>SPELLED</b>    | 0    |
| <b>BURNT</b>      | 1    |
| <b>BURNED</b>     | 0    |

Table 8: Standard and non-standard forms in the HKCSE

Now, as it has been explained in the previous section, the data must be normalised for the sake of comparison. The results are represented in the following table:

|                | <b>BNC</b> | <b>HKCSE</b> | <b>COCA</b> |
|----------------|------------|--------------|-------------|
| <b>GOT</b>     | 92,90      | 130,66       | 105,63      |
| <b>GOTTEN</b>  | 0,10       | 0,55         | 5,77        |
| <b>PROVED</b>  | 6,98       | 0,55         | 2,82        |
| <b>PROVEN</b>  | 0,81       | 0,77         | 2,14        |
| <b>DREAMT</b>  | 0,27       | 0            | 0,18        |
| <b>DREAMED</b> | 0,74       | 0            | 1,02        |
| <b>SPELT</b>   | 0,45       | 0            | 0,03        |
| <b>SPELLED</b> | 0,26       | 0            | 0,47        |
| <b>BURNT</b>   | 1,29       | 0,11         | 0,53        |
| <b>BURNED</b>  | 1,46       | 0            | 2,47        |

Table 9: Normalised frequencies in all three corpora

All of the results obtained from these data and these three corpora will be comparatively discussed in Chapter 5, section 5.3.



## 5. Discussion of the results

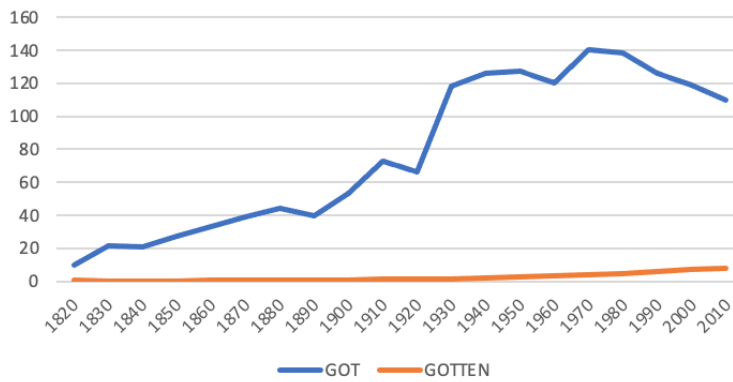
This chapter discusses the results presented in Chapter 4. Therefore, it is divided into three different sections:

- American English. In this first section the data retrieved from COCA in 4.1. are discussed. The data concerning the five selected forms of non-standard past participles are discussed individually in order to answer two of the objectives of this work: (a) to trace the origin of these forms; and (b) to study their development.
- The case of *-ed* suffixes: *dreamed*, *spelled* and *burned*. This section focuses on these three non-standard forms with *-ed* suffixes since they have a high number of characteristics in common.
- The process of Americanization. In this last section, the data concerning the 4.2. section are discussed. On the one hand, it focuses on the comparison of the normalised data retrieved from COCA and BNC, and on the other hand, it deals with the comparison of the data retrieved from COCA and HKCSE. The objective is to compare these three varieties in order to study the development and usage of these non-standard forms, and ultimately, to determine if there has been an actual process of Americanization in the dissemination of the selected forms.

### 5.1. American English

In this first section of this chapter, the data obtained from *the Corpus of Historical American English* (COHA) are discussed.

First of all, the case of the non-standard form *gotten* is analysed. In Table 1 (Chapter 4, section 4.1.1.), the occurrences of the standard form *got*, and the non-standard form *gotten* have been represented. In addition, the results are also represented in the following graph:



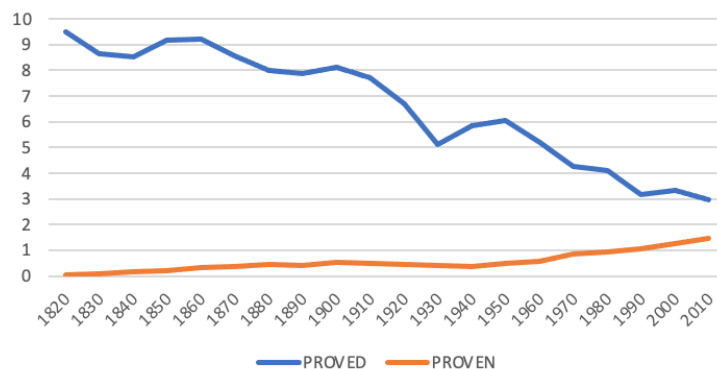
Graph 1: *Got* and *gotten* in the COHA

While *got* had a total of 1551,97 occurrences in COCA, *gotten* only had 43,82 occurrences. Here are some examples of the non-standard form *gotten* retrieved from COHA:

- He might not have *gotten* all he wanted, but he secured what he could (COHA, Rosenbaum, 2011, p. 498).
- “We are interested in appealing to younger people than we've *gotten* before, ” says Mavrikes, describing the company's focus as “ millennial theater, trickling down to Generation Z. ” (COHA, Washington Post, 2019).

The use of both forms has increased over time, but while *got* already had 9,85 occurrences in the decade of 1820-1830, *gotten* only had 0,41 occurrences. This non-standard form is part of Class 5 according to Crystal's distribution. Since this verb is widely used, as mentioned earlier, its standard form persisted and presents a high number of occurrences in AmE. Nevertheless, it can be seen that in the last fifty years (from 1960 to 2010) the number of instances of the non-standard form *gotten* have notably increased, from 3,27 occurrences in 1960 to 7,55 occurrences in 2010, doubling the occurrences. And, in what concerns the standard form *got*, its occurrences decreased in the last decades: in 1990 it amount to 126,03 occurrences, while in 2010 it presented a total of 109,81 occurrences. So, while *gotten* almost doubled its frequency in COHA in the last decades, the standard form *got* decreased.

Table 2 (Chapter 4, section 4.1.2.) represents both standard form *proved*, and non-standard *proven*, comparatively. The results are reflected in the following graph:



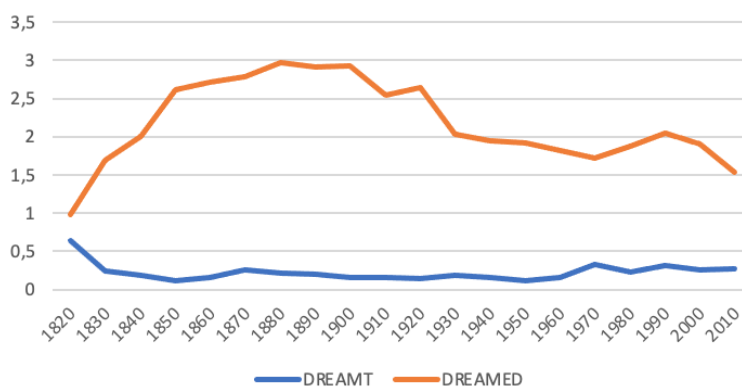
Graph 2: *Proved* and *proven* in the COHA

First of all, here are some examples from COHA of the non-standard form *proven*:

- The book had already *proven* it had an audience when I printed copies for my friends and family, so I went the self-published route to satisfy local demand (COHA, Staheli, 2010, p. 26).
- A second idea is to learn causal models, because they have been *proven* to be more transportable and therefore more robust to changes in the context of decision making (COHA; Dietterich, 2017, p. 3).

At first sight, the standard form has a total number of 132,07 occurrences, while the total number of the non-standard form is 10,83 occurrences. *Proven* had only 0,04 occurrences in 1820 decade, which possibly marks its origin in American English. Its occurrence slowly increased, but not drastically. As far as *proved* is concerned, it also increased over time especially in the first decades of this survey, but its occurrences are different throughout the 20<sup>th</sup> century. It can be said that the forms *gotten* and *proven* have many things in common. Nevertheless, even if the frequency of its non-standard forms has increased, it still cannot be said if there has been a process of Americanization concerning these forms. To do so, they will be compared with other varieties in section 5.3.

Third, the case of the non-standard form *dreamed* is analysed. The data retrieved from COHA show the number of occurrences of this form. The normalised figures are represented in the following graph:



Graph 3: *Dreamt* and *dreamed* in the COHA

According to the normalised data, the standard form *dreamt* has a total of 4,46 occurrences in the corpus, while the non-standard form *dreamed* has a total of 43,57 occurrences. According to Crystal's division by classes, this verb is part of class 2. As he explains, the *-ed* form is more common in American English, which can clearly be seen in the number of total occurrences: *dreamed* has ten times more occurrences than *dreamt* in COHA. But the non-standard forms of *-ed* suffixes might be used as both past participles and preterits. Hereunder some examples retrieved from COHA of both forms are represented:

- The *-ed* preterit form:

- They *dreamed* of capturing Baghdad, dethroning the Abbasids, and uniting the entire Muslim world under their rule (Gray, 2010, p. 16).
- Marina Kamen always *dreamed* of becoming a famous singer, songwriter, and Broadway star (COHA, Knadler, 2007, p.148).

- The *-ed* past participle form:

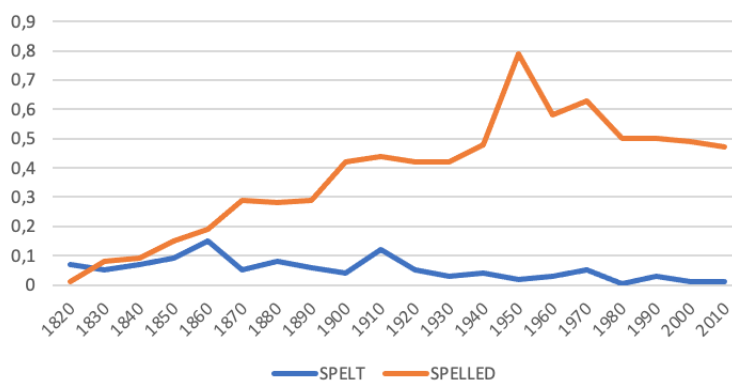
- Faragher certainly hadn't *dreamed* of such things back in the late 1980s, while working as a lifeguard at the Boca Raton beaches to help pay her

way through Florida Atlantic University there (COHA, Carter, 2000, p. 56).

- He had *dreamed* that the screen managed to yelp Fatal Error before the hard drive went to meet its ancestors (COHA, Otoui, 2010, p. 189).

As mentioned, after surveying its uses, this form occurs as both preterit and past participle, being the latter the most frequent. In the 1820-1830 decade, both *dreamt* and *dreamed* had a similar number of occurrences in the corpus, 0,64 and 0,98 respectively, meaning this that they were both used at a similar scale. Nevertheless, from 1830 the non-standard form *dreamed* notably increased, getting up to a maximum of 2,93 occurrences in 1900, while the highest number of *dreamt* was 0,64 occurrences also in 1820. So, Crystal's statement about this form being more common in American English is for now correct. Nevertheless, it will be studied again in section 5.3. where it will be compared with other varieties of English.

Fourth, the case of the non-standard form *spelled* is represented. The results can be observed in the following graph:



Graph 4: *Spelt* and *spelled* in the COHA

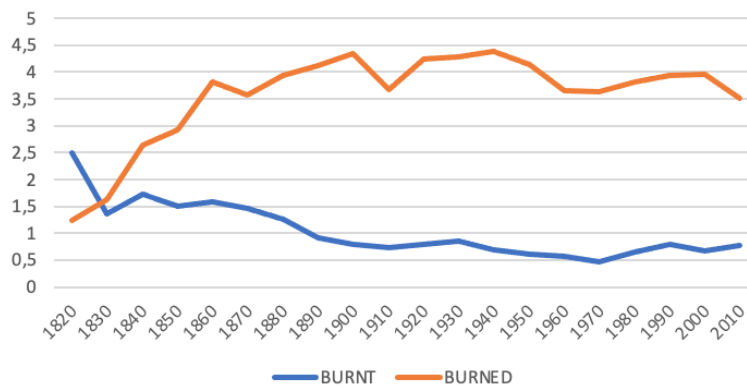
According to the data collected from COHA, the non-standard form *spelled* had a total number of 7,51 occurrences, while the standard form *spelt* had only 0,98 occurrences. This verb, also being part of Crystal's class 2, has similar features as

*dreamed*. The origin of this non-standard form goes back to 1820, when it only had 0,01 occurrence, while it reached its highest frequency in 1950, with 0,79 occurrences. In addition, the non-standard forms of *-ed* suffixes have been used as both past participles and preterits in COHA. Below are some examples retrieved from the corpus of both forms:

- The *-ed* preterit form:
  - Even before Albert *spelled* out her argument, the magazine had clustered the stars into two distinct categories, to be recognized at a glance (COHA, Keating, 2017, p. 105).
  - "Hey, we have a Teresa here, *spelled* the same way. " The girl herself sometimes came to her table (Alcala, 2000, p. 11).
- The *-ed* past participle form:
  - Only a food fight in the House cafeteria could have better *spelled* out the end of DeLay's once vaunted and feared iron discipline (COHA, Verini, 2006, p. 17).
  - But things have not been *spelled* out as to just what it is that is going to be okay (COHA, Brooks, 1990, p. 22).

Overall, the use of this form is more common in past participle tenses in the corpus. It can be said that the non-standard *spelled* is far more frequent in American English than the standard form *spelt*. Nevertheless, more details on the process of Americanization of this and the other forms will be addressed in section 5.3.

And finally, the case of the non-standard form *burned* is discussed. The normalised data obtained from COHA are represented in the following graph:



Graph 5: *Burnt* and *burned* in the COHA

The standard form *burnt* had a total of 20,7 occurrences in COHA, while the non-standard form *burned* had a total of 71,4 occurrences. Examples follow concerning its preterit and past participles forms that have been retrieved from COHA for this study:

- The *-ed* preterit form:
  - I was the last person in the church the night it *burned* down (COHA, Kashner, 2010, p. 6).
  - This was the same Alonzo Mourning who had just donated \$ 50,000 worth of suits to a tall man whose house *burned* down, but no one in New York City knew (COHA, Friend, 2000, p. 130).
- The *-ed* past participle form:
  - Our seas have been plundered by King George III, our coasts have been ravished, our towns have been *burned*, and the lives of our people have been destroyed (COHA, Kovach, 2010, p. 28)
  - The firefighters shrug and point to homes they did save, including pine-draped bungalows that by all rights should have *burned* (COHA, Flashover, 2001).

It can be said that the origin of *burned* goes back in time, probably even earlier than 1820, since in this period it already presented 1,23 occurrences in a corpus of 100,000 words. Both *burnt* and *burned* have presented a significant number of occurrences in COHA, but while the occurrences of *burnt* never increased drastically, the occurrences of *burned* did. It started with only 1,23 occurrences in 1820, and presented

more than 4 occurrences in some decades of the 20<sup>th</sup> century. Therefore, it can be concluded that the popularity of use of this form especially grew in the 20<sup>th</sup> century, while the standard form almost always maintained its frequencies.

Overall, it must be said that the first two forms *gotten* and *proven*, have many aspects in common, while the last three forms *dreamed*, *spelled*, and *burned*, have other ones in common according to the data retrieved from COHA. In what concerns *gotten* and *proven*, the standard forms of these verbs (*got* and *proved*) have presented far more occurrences than the non-standard ones in AmE. Their standard forms are much more consolidated in this variety than the non-standard ones. Nevertheless, in the near past, their non-standard forms started increasing, which means that maybe in a future they will be more widely used. The study of their development is clearly an interesting and challenging task for researches interested in this field. The use of the non-standard forms may have different explanations: it can happen because the speakers are unsure of what form to use or which one is the correct form to use, as J. Cheshire (1993) explained; or it can happen because the non-standard form seems easier to pronounce and write for the speakers, and seems more natural to them, as L. Anderwald (2011) explained. Nonetheless, what happens with *dreamed*, *spelled* and *burned* is completely the opposite. The non-standard form of these verbs happened to be much more frequent than the standard one in COHA. So, at first sight, Crystal's statement can be affirmed in this research. Nevertheless, in order to be sure, the data obtained from the American English variety will be compared to the data obtained from the British English and Hong Kong English varieties. This way, some conclusions will be made considering the process of Americanization of these non-standard forms (in section 5.3.). Nevertheless, the next section of this chapter will focus on these last three non-standard forms.



## 5.2. The cases of *-ed* suffixes: *dreamed*, *spelled* and *burned*

This section focuses on the three *-ed* forms that have been studied in this work: *dreamed*, *spelled* and *burned*. When looking them up in the *Cambridge Dictionary*, they appear as typical forms in American English, while the standard forms *dreamt*, *spelt*, and *burnt*, are of British use. In addition, as it has been mentioned in Chapter 2, Crystal (1995) divided the irregular verbs into seven classes. The verbs that have alternative forms: the *-ed* suffix (common in American English), and the *-t* suffix (common in British English), are verbs from Class 2 (p. 204). This is the case of the three non-standard forms under scrutiny: *dreamed*, *spelled*, and *burned*. Crystal (1995) explains two important things that have to be considered about these verbs and their two forms. First, he explains that the use of one form or the other has to do with the action's duration. Crystal (1995) states that:

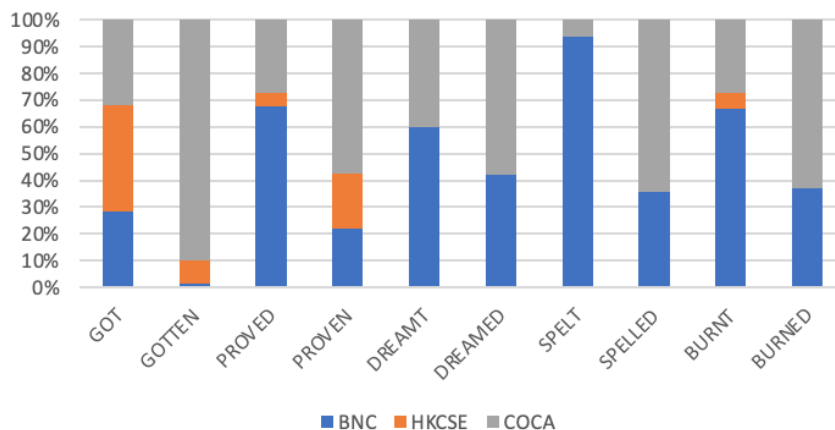
The close comparison of examples suggests that *the -ed* form may be more likely when the duration of an action is being emphasized. Something which has happened once, which was taken up very little time, or which focuses on the result of a process rather than on the process itself may be more likely to attract the *-t* ending. (p. 204).

So, first of all the duration of the action must be considered. Second, he states that it may happen randomly “because the verb or context does not motivate the drawing of such semantic distinctions”, and “there are some collocations which permit little or no variation” (Crystal, 1995, p. 204). It is true that there are different collocations that do not allow variation, as for example *spelled out*, that has been represented as an example in section 5.1. Nevertheless, these are not the only things to keep in mind. There are also other aspects that influence the speaker when choosing one form over the other. As it has been explained in this work, the speaker can chose one form over the other due to the lack of knowledge of the correct form or because of the more ‘natural’ way than one form sounds over the other. Moreover, M. Levin (2009) states in his work that: “The variation

in preterite and past participle forms of verbs such as burn, dream, leap and spell is one of the most cited morphological differences between American English (AmE) and British English (BrE)” (p. 60). Levin explains that these forms attract more and more researchers since it is definitely one of the most popular differences between these varieties. Nonetheless, in order to know exactly what happened with the use of this form, and as Levin said, they must be compared with other varieties.

### 5.3. The process of Americanization

In this section the results regarding the different varieties of English are discussed. The normalised data that have been calculated earlier from the *Corpus of Contemporary American English* (COCA) are compared with the data obtained from the *British National Corpus* (BNC) and the *Hong Kong Corpus of Spoken English* (HKCSE). This section is concerned with one of the objectives of this work as well as its research question: is there an actual process of Americanization of non-standard past participles? In the following graph, their distribution across the varieties is represented:



Graph 6: Normalised data from the corpora

- *Got* and *gotten*: The highest number of occurrences of *got* is observed in HKCSE with 130,66 occurrences. Then, COCA had a total of 105,63 occurrences, followed by British English with 92,9 occurrences. But, as far as the non-standard

form is concerned, *gotten* had clearly the highest number of occurrences in COCA when compared with the three other corpora: 5,77 occurrences. Even though it is a much smaller number than that of the standard form, it is still significantly higher than the numbers of the other varieties of English that have been studied: 0,55 in HKE and 0,1 in BrE. Therefore, it can be concluded that *gotten* is clearly more frequent in AmE if compared with the other varieties, and therefore, that it can be considered to be representative of this variety.

- *Proved* and *proven*: In the case of the standard form, *proved*, it presented the highest distribution in BNC with a total of 6,98 occurrences while 2,82 occurrences were retrieved from COCA, and only 0,55 occurrences from HKCSE. Regarding the non-standard form, *proven*, both BNC and HKCSE corpora presented a smaller number of occurrences, with just 0,81 and 0,77, respectively, while 2,14 occurrences were retrieved from COCA. As explained earlier, even though the standard form is more widely used than the non-standard one in AmE, its use started to increase notably in the last decades (at the end of the 20<sup>th</sup> century and during the 21<sup>th</sup> century). Overall, the form *proven* is the most frequent one in this variety, but not that frequent to be considered exclusively representative of AmE.
- *Dreamt* and *dreamed*: *Dreamt* had the highest number of occurrences in British English (0,27 occurrences), as Crystal (1995) stated, since it is a commonly used form in this variety. HKCSE had no occurrences of neither standard nor non-standard form, while COCA had a total of 0,18 occurrences. Nevertheless, in what concerns the non-standard form *dreamed*, COCA had 1,02 occurrences, and BNC 0,74 occurrences. The AmE variety presents the highest number of occurrences in the three corpora, but the difference between it and the BrE variety is still not significant. Due to this data, this form cannot be considered to be representative of AmE.
- *Spelt* and *spelled*: As it happened with *dreamt*, *spelt* had the highest number of occurrences (0,74) in BrE. These verbs that belong to class 2 of Crystal's distribution have the same characteristics. HKCSE also had no occurrences of neither of these forms, while AmE once again had the highest number of occurrences of the non-standard form in *-ed*, *spelled*, with a total of 0,47 in comparison with the 0,26 occurrences which were retrieved from BNC. Nevertheless, the difference in this case between BrE and AmE is not that

significant, and that is why it cannot be completely affirmed that this form is representative of AmE only.

- *Burnt* and *burned*: Regarding *burnt* it presents the highest number of occurrences in BNC, with 1,29 occurrences, followed by 0,53 occurrences in COCA, and just 0,11 occurrences in HKCSE. As far as the non-standard form *burned* is concerned, it presents a total of 2,47 occurrences in the AmE variety, no occurrences in the HKE variety, and 1,46 occurrences in the BrE variety. Even though this form has the highest number of occurrences in AmE, its numbers are not high enough for it to be representative of this variety yet.

Overall, it can be said that on the one hand, all of the standard forms that have been studied in these three varieties have been more frequent in both BrE and HKE than in AmE. While on the other hand, it can be seen that the non-standard forms have proved to be popular in AmE. In all of these cases, the data retrieved from COCA present the highest number of occurrences. Nevertheless, the use of these forms in AmE is not much higher when compared to the other varieties, especially in the case of BrE. Therefore, it can be said that some non-standard forms such as *gotten* show a clear process of Americanization, but forms such as *proven*, *dreamed*, *spelled*, and *burned* do not present enough evidence to claim for Americanization nor to be considered exclusively representative of this variety.

## 6. Conclusions

This work analyses the distribution of non-standard forms of past participles in American English diachronically with the help of a representative corpus of this variety: the *Corpus of Historical American English* (COHA). It starts with Crystal's statement that English is a global language. The history of the diffusion of the English language, going from the influence of the British Empire to the economic supremacy of the United States in the twentieth century leads to their present-day's leading force in the technological industry. Nowadays, the U.S. has a great impact on the English language and therefore the present dissertation delves into some particular features of AmE. Having a look into corpus-based studies in language, the definition, origins and future of Corpus Linguistics, and having addressed the topic of American English paying special attention to the concept of Americanization, helped establishing the theoretical framework and therefore, the basis of this research.

In this work, five different forms of non-standard past participles have been studied in a period of a hundred and ninety years, from 1820 to 2010. To do so, a quantitative approach was used. Moreover, it has been possible to view the words in context, with the KWIC option, so that each result could be viewed as a whole sentence or text. First of all, the data from COHA of both standard and non-standard forms were compared between each other: *got* and *gotten*, *proved* and *proven*, *dreamt* and *dreamed*, *spelt* and *spelled*, and *burnt* and *burned*. Thanks to this, many conclusions have been made. In what concerns *gotten* and *proven*, its standard forms continue to have a bigger number of occurrences than the non-standard ones. Nevertheless, in the last years, especially in the second half of the 20<sup>th</sup> century and during the 21<sup>st</sup> century, the frequency of these forms have notably increased. Therefore, some assumptions on their possible future development have been made. Regarding *dreamed*, *spelled* and *burned*, some

differences were spotted. These forms proved to be very popular in AmE, since they presented a much bigger number of occurrences in the corpus, and therefore, Crystal's theory was confirmed. By analyzing these forms in COHA, two of the objectives of this work have been accomplished: (a) to trace the origin of these forms; and (b) to analyze their development.

Afterwards, the data obtained from COCA have been analyzed and compared with the data obtained from the British English variety, through the *British National Corpus* (BNC), and to the data obtained from the Hong Kong English variety, through the *Hong Kong Corpus of Spoken English* (HKCSE). After the normalisation of the data (the corpora that were used for extracting the data had different lengths), a comparison between the varieties has been possible. By doing this, the last objective and research question of this work have been answered. In the case of *gotten*, it proves to be representative of AmE concluding that there has clearly been an actual process of Americanization. But, in the case of *proved*, *dreamed*, *spelled*, and *burned*, the so-called process of Americanization cannot be claimed in the light of their constrained occurrence in the corpora. Overall, it can be confirmed that all of these non-standard forms, even though to a different extent, are influenced by Americanization. The reasons which may eventually justify the use of one or another form are just a matter of hypothesis. Even though the reason is not exactly known, the speakers of a language might opt for a specific form due to its more 'natural' sound, as Anderwald (2011) stated, or due to their lack of knowledge since many do not speak a regulated language, as J. Cheshire (1993) explained. It can even happen randomly or the choice can be influenced by the duration of an action, as Crystal (1995) explain. Therefore, this dissertation definitely invites to further research in what concerns irregular verbs or forms in English, as Crystal (1995) stated:

An irregular lexical verb is one where some of the forms are unpredictable. There are thousands of regular verbs in Modern English, but less than 300 irregular ones. The irregular forms are the surviving members of the highly developed system of ‘strong’ verb classes found in Old English”. (p. 204).

He recognizes the importance of these verbs and their non-standard forms, and therefore we invite researches to continue investigating different phenomena in this field. Overall, it has been a challenging task to survey a big amount of data retrieved from different corpora, but with a proper use of Corpus Linguistics and working with well-planned and distributed corpora, it was made possible.

## References

- Anderwald, L. (2011). Are non-standard dialects more 'natural' than the standard? A test case from English verb morphology. *Journal of Linguistics*, 47(2), 251-274.  
Retrieved June 10, 2021, from <http://www.jstor.org/stable/41261753>
- Cambridge dictionary of American English (2nd ed.). (2008). Cambridge University Press.
- Cheng W, Greaves C, Warren M (2005). The creation of prosodically transcribed intercultural corpus: The Hong Kong Corpus of Spoken English (prosodic), *ICAME Journal*, vol. 29 (pg. 47-68), April 2005.
- Cheshire, J. (2012). Standardization and the English irregular verbs. In D. Stein & I. Tieken-Boon van Ostade (Ed.), *Towards a Standard English* (pp. 115-134). Berlin, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110864281.115>
- Crystal, D. (2003). *English as a global language* (Second edition.). Cambridge University Press.
- Crystal, D. (1995). *The Cambridge encyclopedia of the English language*. Cambridge University Press.
- Davies, Mark. *The Corpus of Contemporary American English (COCA): 560 million words, (1990-present. 2008)*. <https://corpus.byu.edu/coca/>.
- Davies, Mark. *The Corpus of Historical American English (COHA): 400 million words, (1910-2009. 2015)*. <https://www.english-corpora.org/coha/> .
- Gonçalves B, Loureiro-Porto L, Ramasco JJ, Sánchez D (2018) Mapping the Americanization of English in space and time. *PLoS ONE* 13(5): e0197741. <https://doi.org/10.1371/journal.pone.0197741>



- Hernando Carnicero, Carmen. (2009). A Corpus of American and British English: A Case Study of Slang. En: *Anglogermanica Online*; Revista electrónica periódica de filología alemana e inglesa, 2009-2010, No. 7: 3-18.
- Levin, M. (2009). The formation of the preterite and the past participle. In G. Rohdenburg & J. Schlüter (Eds.), *One Language, Two Grammars?: Differences between British and American English* (Studies in English Language, pp. 60-85). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511551970.004
- Newman, J., Baayen, R. H., & Rice, S. (2011). *Corpus-based studies in language use, language learning, and language documentation*.
- Pereira Gomes Da Costa, H. (2019). Multilingual data collection for multiple corpus-based approaches to translation and interpretation. *UMA Editorial*.
- Scheibman, J. (2002). *Point of View and Grammar. Structural patterns of subjectivity in American English conversation*. John Benjamins Publishing Company.
- The British National Corpus*, version 3 BNC XML Edition (2007). Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium.  
URL: <http://www.natcorp.ox.ac.uk/>
- Viana, V., Zyngier, S., & Barnbrook, G. (2011). *Perspectives on corpus linguistics*. J. Benjamins Pub.
- Wikström, Eric. (2013) *Non-Standard –ed Forms of Selected Irregular Verbs: A Corpus-based Study of Present-day American English*. Göteborgs Universitet.