Excellent Language & Linguistics Dissertation Repository (ELLDR), Newcastle University, Vol. 3, 2023-24.

TRACING THE USE AND VARIETY OF NEW WORD SIGNALLING TECHNIQUES IN THE ENGLISH LANGUAGE FROM 1610 TO 1960

Jennifer Ann Hargreaves
English Language, 2024

Abstract: The overall aim of this dissertation is to analyse the use of 'discriminants' that identify neologisms in a text from the seventeenth to the twentieth century. Studies of neologism discriminants have focused on the efficiency of compiling new word lists, leaving a gap in the field regarding their historical use, including when certain discriminants began to be used in texts, and whether there are any discriminants that have dropped out of use in texts. In this dissertation I provide an in-depth assessment on the marking of neologisms, building upon previous research of the extraction of neologisms through discriminants presented in Paryzek (2008). Using the *Oxford English Dictionary* (OED) as the primary data source, I offer further insight into the use of the discriminants outlined in Paryzek (2008) and present discriminants not previously identified in neologism research that have been used at different points in time, found through the studying of first attestation quotes. The frequencies of use of the discriminants are tracked, alongside the most common discriminant type for each year and which word types are most frequently signalled.

Keywords: neologisms, new word signalling, discriminants, lexical discriminants, borrowing, compound, blend, affixation, loan blend, quotation marks, italics

Supervisor(s): Dr William van der Wurff

TRACING THE USE AND VARIETY OF NEW WORD SIGNALLING TECHNIQUES IN THE ENGLISH LANGUAGE FROM 1610 TO 1960

1 INTRODUCTION

There are multiple methods of highlighting new words to audiences in the English language. Termed "discriminants" by Paryzek (2008: 165), quotation marks and specific lexical indicators, such as *called* and *termed*, are utilised in the automatic retrieval of neologisms from texts on the Internet.

However, these discriminants are only used in research as a tool to collect data on neologisms, and very little attention has been paid to the discriminants themselves. This research therefore focuses on the historical highlighting of new words to audiences, tracking the use of discriminants that appear alongside neologisms in their purported first use in English at fifty-year intervals, using data gathered from the *Oxford English Dictionary* (*OED*). The range of discriminants used is an important area of focus in this research, going beyond the few suggestions of suitable lexical markers presented by Paryzek (2008), as well as the tracking of the use of these discriminants over time, with the use of data from 1610 to 1960.

Taking into consideration the work of Durkin (2014), who details a large increase in the total number of new words recorded in the *OED* from the 1550 to 1700, and from 1800 until the end of the twentieth century, I make the prediction that there will be a peak in overall discriminant use in 1610 and 1660, and again in 1910 and 1960, after a dip in use in 1710 and 1760. I also predict that that there will be a steady increase in the range of lexical markers used to highlight a new word to audiences, based on the work of Baayen et al. (2017) which noted that over the previous two hundred years there has been a gradual growth of the variety of different words in use in English.

This study starts with a look into neologisms, how they can be perceived by writers and audiences and how they can be formed, followed by an outline of the research carried out in Paryzek (2008) and the conclusions in that research that inspired this study. The use of discriminants by other researchers in neologism research is also mentioned in this section. This is then followed by my research focuses for this study, where I emphasise my interest in

the range of lexical discriminants used, the frequency of different discriminant types, and how methods of new word formation may correlate with the discriminant type used to signal it. I predict that quotation marks will be the most frequent discriminant throughout the time period I have selected, that the range of lexical discriminants will increase year on year, and that the most frequent lexical discriminant will be called. I then outline my data collection method in Section **3**, selecting a sample of 500 words for each of the eight years I have selected as part of the study. In Section **4**, I provide an overview of discriminant use over the eight selected years, finding that the use of discriminants does largely increase from the nineteenth century, and that they are helpful resources in identifying neologisms. It is found that lexical discriminants, quotation marks and italics are the three types of discriminants used, and that lexical discriminants and italics are initially the most frequent, before quotation marks become the most frequent discriminant type from 1860 onwards. I also provide some insight into the discriminant frequency for each individual year, analysing which word types are most frequently used alongside the three discriminant types. In the discussion of Section **5,** I outline the multiple areas of research this study opens up in relation to discriminants, including further research into how discriminants have been used historically. Finally, I conclude that if a larger study of the same nature was to be conducted, it would be beneficial to ascertain whether there is a correlation between word type and discriminant type, and whether the text type can influence the type of discriminant used.

2 BACKGROUND

This section starts with an overview of neologisms, including their potential definitions, followed by an overview of how neologisms are formed, as I believe that both of these things will play a part in how discriminants are used in texts. Neologisms can either be signalled because they are completely new, or because they have only been used occasionally and require some kind of explanation on part of the writer. As this study is not able to separate out these types of neologisms, I have included the definitions for both, and both will be considered in this study. Following this is an outline of the research of Paryzek (2008) who tested the productivity of a select few discriminants, which inspired this study that focuses on the discriminants themselves as opposed to the neologisms that they are signalling. After some further discussion of discriminants used in automated neologism retrieval studies, I present the focuses of this study and what I am anticipating to find.

2.1 NEOLOGISM DEFINITIONS

A neologism can be defined as "an item newly introduced into the lexicon of a language" (Anderson and Malmkjær 2006: 601). The monitoring of newly introduced words provides the opportunity to carefully study the processes of ongoing change (Kerremans et al. 2011), and to study the societal trends that neologisms reflect (Ayto 1996).

Given the focus of this dissertation on the highlighting of new words in texts, one question that arises is: What kinds of new words get highlighted by discriminants? To answer this question, it is important to consider what would lead a lexical item to being considered a new word. In the literature there are several approaches to the definition of new words that have been discussed, which are summarised in Ološtiak and Rešovská (2011).

Not all of the frameworks outlined here will be considered in the results analysis for this study, but they are important for considering the various ways in which lexical items can be regarded as 'new'. These frameworks include one that views a neologism as exclusively a lexeme that has led to a significant change in a lexical system, named the 'language-oriented perspective' by Ološtiak and Rešovská (2021: 304), and one that only considers lexemes occurring in utterances to be genuine neologisms, named the 'speech-oriented perspective' (Ološtiak and Rešovská 2021: 304).

One framework for the analysis of neologisms that is important for this study is presented by Ološtiak and Rešovská (2021: 304), who outline the 'user-oriented perspective'. This focuses on the language user's perception of a lexeme's newness, and, when applied to this research in particular, can be used to approximate a writer's knowledge regarding a word and how likely the audience is to be able to identify and understand it. This is particularly important when considering data from the *OED* as there is a chance that a lexeme was already in limited use before the date of first attestation provided. As it can be difficult to determine a writer's assessment of a reader's knowledge of a word, or the writer's understanding of their audience, all cases discussed in this research will be considered as part of a user-oriented perspective, with the assumption that words recorded by the *OED* were not in general use at the date of their purported first attestation, and that any discriminants used in relation to the word were considered useful by the author as they might have felt that the word was largely, or entirely, unknown, and potentially difficult to understand.

Whilst this research is only considering physical texts, it is evident that the individual users of a language are an extremely important aspect of neologism analysis.

2.2 NEW WORD FORMATION METHODS

It is also crucial to consider the various methods of new word formation that are utilised in the English language, and how the highlighting of a new word could potentially vary depending upon the way in which a new word is created.

One of the most common ways of adding words to the English language is through compounding (Lieber 2005), which involves the combination of two independently meaningful roots in order to form a new word (Harley 2006). Such roots can be nouns, verbs, and adverbs. The productivity of this word formation method is very high, with noun-noun compounds being completely productive in English (Plag 2003: 145). Despite being highly productive, compounds can still be less than fully transparent, as the meaning of the entire new word cannot always be deduced by considering the individual meanings of its individual components. As demonstrated in the example provided by Harley (2006: 100), the compounds *nurse shoes* refers to shoes that are made for nurses to wear, but *alligator shoes* does not refer to shoes that are made for alligators to wear, and is instead referencing shoes that are made from the skin of alligators. For any specific noun-noun compound, a language user would need to understand what semantic relation between the two nouns is intended, despite there being no overt indicator of this difference in the combined word itself. This lack of transparency for some compounds could lead to the use of highlighting techniques by an author to indicate to their audience that a new word with a potentially ambiguous meaning is being introduced and that they may need to devote some effort to establishing its intended meaning.

Another highly productive method of word formation in English is derivation (Lieber 2005). This involves the creation of a new word from an existing one through the use of non-inflectional affixes. The meaning of the new word can be largely transparent to a reader, as long as they understand the meaning of the individual prefixes and/or suffixes and the root word, and how these combined elements work together.

Slightly less prevalent in word formation in English is the method of clipping, which reduces a multisyllabic word to one or two syllables (Harley 2006: 95). Harley (2006) notes that it is common for a word to be clipped because it has become common usage, and language users no longer feel the need to use the full version to express the concept. An example of clipping includes the reduction of *gymnasium* to *gym*, or the reduction of *motorbike* to *bike*. This is one of the least transparent word formation methods, as it can be difficult to identify the word that has been reduced, and there is also the possibility that a clipped form could be a homonym of a separate word that already exists.

A blend is also a method of clipping, in which two words are clipped and combined together to create a new word, where the meaning is connected to the meaning of the original words (Harley 2006: 101). Blends are suggested to be some of the new words that language users are most aware of, most likely because they were created purposefully (Harley 2006), but they are also amongst the least transparent because of the meaning relying on knowledge of the full-length words. For example, understanding *smog* would be very difficult if you weren't aware that the two separate words were *smoke* and *fog*.

The most extreme kind of clipping is an initialism, which includes acronyms and abbreviations, and involves using the initial letters of content words in a phrase to stand in for the whole phrase (Harley 2006: 96). This is a method of word formation that has been around for a relatively long time in English but gained traction in the latter half of the twentieth century (Harley 2006: 96). Acronyms are a collection of initials that are pronounced as a single phonological word, such as *laser* (light amplification by stimulated emission of radiation) whereas abbreviations are a collection of initials where the letters are read out individually, such as FBI (Federal Bureau of Investigation). This is probably among the least transparent of all new word formation methods, as it would require a language user to know exactly what source word each of the letters in the new word corresponded to.

Finally, there is the process of 'borrowing', in which one language either wholly or partly replicates a linguistic feature from another language (Durkin 2014: 1). Words can have both a form and a meaning, and cases where it is the form that has been borrowed into English are known as "loanwords" (Durkin 2014: 3). In English, loanwords make up a large proportion of the words in any large dictionary and they are a key component in the language of everyday communication, with some loanwords even being found among the "most basic vocabulary of English" (Durkin 2014: 4). Loanwords that borrow the form of words from a donor language whilst also incorporating elements of the borrowing language are known as "loan blends" (Durkin 2014: 9).

It is important to note that as there was a large expansion in the number of loanwords in English during the nineteenth century and beyond, there was also an increase in new words entering the English language as a whole. There is a large increase in the overall number of new words of all types recorded in the *OED* from 1550 to 1700, with a notable drop throughout the eighteenth century, before an even larger peak of new words being introduced during the nineteenth and twentieth century. As Baayen et al. (2017) notes, societal changes lead to changes in language, and the rapid developments in all areas of society have led to an increase in the number of different words in use in American English over the past two

centuries. Whilst this can be attributed in part to brand new technological and scientific inventions, there is also an issue of an increased demand on language (Baayen et al., 2017: 19); there is a greater array of concepts and themes known to language users as information becomes more widely accessible, and these concepts often require new words to be created in order to be appropriately named and identified. Using government registration data on personal names Baayen et al. (2017: 3) found that there was an exponential increase in personal name vocabulary from the end of the second world war. It does not seem implausible to suggest that a similar peak of new words from the twentieth century could be recorded in the *OED*, or that there could be a wider range of highlighting techniques used for these words.

2.3 PARYZEK (2008)

The retrieval of neologisms from texts has, until recent decades, traditionally been carried out manually. This manual method would entail the reading of a text and simultaneously noting any unknown lexical units that are encountered in order to check these units against multiple dictionaries and create a final list of neologisms. Whilst highly accurate, the manual method is considered uneconomical in the face of the vast quantities of material presented by modern texts. Therefore, an automated neologism retrieval technique would appear to be the logical next step. This is the precise focus of Paryzek's (2008) research, with the comparison of automatic excerption techniques to test their suitability for neologism retrieval.

As mathematical approaches cannot be used to identify the presence of a neologism in a corpus or database, Paryzek pulls upon previous research by Chlebda (1991), which highlighted that Polish neologisms occur within quotation marks or after certain phrases, and tests whether a similar phenomenon can be found in the English language.

The explorational of the retrieval of neologisms in Paryzek's study focuses on the graphical and lexical items in a text, and whether these "easily retrievable entities" that are within the same "neighbourhood" as the neologisms occur in this environment to a greater extent than they appear elsewhere in the text (2008: 165). These retrievable entities are referred to by Paryzek as "discriminants" and encompass both the lexical items and the punctuation items (2008: 165).

Lexical discriminants are described by Paryzek as phrases that usually, or with a greater frequency, occur before a neologism in a text, indicating to the reader that a word is

new, or helping to define it. As part of this research, a limited number of discriminants were chosen to be tested as retrieval tools:

. *termed*

. *called* (and *so-called*)

. *known as*

. *defined as*

Punctuation discriminants were also included in the form of single and double quotation marks as they were expected to surround phrases that were likely to contain a neologism.

Using automatic software, Paryzek investigated whether the lexical and punctuation discriminants did actually signal that neologisms were present in a text. This was achieved with the use of a complete collection of articles published in the scientific journal *Nature* over a nine-year period from 1997 until 2005, with the software identifying any instance in which the lexical or punctuation discriminants were used in a sentence.

The results of this experiment showed that the single quote discriminant was the most productive of all the discriminants, with the highest ratio of the number of neologisms extracted to the total number of words analysed. In regard to the lexical discriminants, it was concluded that they were productive, and useful in identifying neologisms in a text, with *termed* and *called* being more productive than *known as* and *defined as*.

2.4 DISCRIMINANTS IN OTHER RESEARCH

Although discriminants are not the key focus of neologism research, there have been various references to them in the work of other linguists. Whilst there is variation in how these lexical items are named by the linguists, there is a range of research that indicates these items are an important aspect of neologism identification and retrieval.

One such example is in the work of O'Donovan and O'Neill (2008), who outlined a systematic approach to selecting neologisms that would be included in a monolingual dictionary. This work utilised linguistic patterns that marked lexical novelty, such as punctuation marks, in order to identify potential neologisms that could be compiled and verified at a later date, indicating that discriminants could be a common feature of multiple languages.

Abel and Stemle (2018) also used discriminants to semi-automatically extract neologism candidates from a non-English language, focusing on the German standard variety

that is used in the Northern Italian province of South Tyrol. Lexical indicators were also suggested to be of use in retrieving neologisms from Japanese texts, who also applied machine learning techniques in order to identify any constructions that can be associated to new words in a language (Breen 2009).

The NeoCrawler tool designed by Kerremans, Stegmayr, and Schmid (2011) uses Google search pages in order to add neologisms to a database. Automated analyses of the search pages generated result in compiled tokens, and concordance lines for each token. This tool detects "novelty markers" that include phrases, such as "so-called" and quotation marks (Kerremans et al. 2011: 74). NeoCrawler also searches for longer strings such as "came up with a/the (new) term/word", "made up a/the (new) term/word", "invented a/the (new) term/word", and "coined a/the (new term/word" in order to find neologisms in texts (Kerremans et al. 2011: 78).

Interestingly, Smyk-Bhattacharjee (2006) paid a little more attention to discriminants and their use by authors in her work on the interpretation of novel words. Referring to this signalling as "distancing strategies", Smyk-Bhattacharjee notes that an author can use these techniques in order to mark the word as new to their lexicon, or as one that they have yet to accept as a word (2006: 32). Specific examples included of these techniques include phrases that an author could use to directly state a concept is new, such as "new field", or phrases that allow for impersonal distancing from the neologism, such as "sometimes labelled" and "what is referred to as…" (Smyk-Bhattacharjee 2006: 32).

Evidently, the lexical items that Paryzek identifies as discriminants are useful in the retrieval of neologisms, and this usefulness has been identified in multiple languages. However, for all of this mention of 'markers of lexical novelty' and 'distancing strategies', there is very little focus on the items themselves. I feel that this is a significant gap in research that could be explored – what is the range of lexical items used to indicate a neologism in a text? Have these items changed in frequency over time? Are certain lexical items more frequently found with certain types of neologisms?

Therefore, this research will investigate these questions, in the hopes of being able to track how neologisms have been signalled over a significant time period, and whether these special lexical items, particularly the lexical phrases, have followed the same trends as new words as outlined in Durkin (2014). I believe that this area of research could provide some insight into language change, and potentially provide some insight into how authors interact with new words and their own audience.

2.5 RESEARCH QUESTION

This study will focus on tracking the use of discriminants in English language texts over a 350-year period. I intend to determine the range of discriminants that have been used to signal a new word, and whether the frequency of particular discriminants have changed over time. This will differ from other research involving discriminants, as only a select few discriminants have been investigated in neologism retrieval, in order to calculate their overall yield. As my focus is on the discriminants themselves and not the words they are signalling, this study hopes to find numerous discriminants that have not been identified in any previous research. I also hope to determine whether there are any trends between the type of new word being introduced and the discriminant that is used to signal it, as this is a very big gap in the research, and the discovery of any trends could be beneficial to the future of neologism retrieval and research, as well as any future research on the discriminants themselves.

The frequency of discriminants will be compared to Paryzek (2008), to determine whether punctuation discriminants are more productive than lexical discriminants, and to determine whether *termed* and *called* remain the most productive and frequent of lexical discriminants when there is a wider range of data used as a source.

This study will also investigate whether there is any pattern between the type of a word and the discriminant that is used to signal it, as I believe that the differing transparency of word-formation techniques could change how they are introduced to an audience in a text.

Overall, I predict that punctuation discriminants will be more frequent than lexical discriminants throughout the entire time period that this study spans across, and that *termed* and *called* will be the most frequent lexical discriminants in the twentieth century. I also predict that the range of lexical discriminants in use will increase from the seventeenth century to the twentieth century, but that there will be a decrease in the overall frequency of any type of discriminant during the eighteenth century, due to the dip in new-word entries to the *OED* as noted by Durkin (2014).

I also expect that loanwords will be the most frequent of word types found in this study, and that they will most frequently appear alongside lexical discriminants.

3. DATA COLLECTION AND METHODOLOGY

This study will be using data manually collected through the Oxford English Dictionary website in order to identify and track the range of lexical items used to identify neologisms.

For the purpose of this study, I will be referring to these lexical items as 'discriminants', either lexical or otherwise, as in Paryzek (2008).

The data used in this study will be acquired from the first attested quote for words, as provided by the *OED*. This is very important, as these quotes will be where I identify any discriminants that were used by the author of the text, and, crucially, these quotes will be dated, indicating when the words were first used or introduced to the English language. Therefore, I can use the *OED* to note when a word was introduced, and whether this introduction included any discriminant. This will allow me to analyse and compare how new words have been signalled in different years.

I have chosen to start looking at words from the year 1610. This will involve searching for words that were first attested in 1610 and reading each provided quote to find any discriminants. Any discriminants will be noted, alongside the 'new word', to allow for analysis. This process will be repeated at 50-year intervals, ending at the year 1960. I believe that this will provide a large amount of data that will be able to be compared, allowing for an insight into whether discriminants have changed in their frequency over the 350-year period.

This method of data collection will allow me to list any discriminant that I find in a spreadsheet, which will then provide the opportunity to find any discriminants that weren't identified in previous neologism retrieval research. However, as the data is being compiled manually, I have decided to look at 500 words per year, unless there is any year in which there is a total number of dictionary entries totalling less than 500, in which case all of the available entries will be analysed. This means that this study will look at approximately 4000 words and their relevant quotations. I believe that this will yield a sufficient amount of data for analysis, and that 500 words per 50-year interval will allow me to accurately determine whether there is a range of discriminants, if any at all, even if not every word provided by the *OED* is analysed. The token frequency of the words that were signalled by some type of discriminant will be collected for each 50-year interval. This will then be converted into a value per thousand words, in order to normalize the frequency scores.

Alongside the word, its date of first attestation, and any discriminant I was able to find in the provided quotation, I will also be noting the type of new-word formation that the *OED* provides in the etymological information for each word entry, and the type of text the first attestation was found in. The multitude of information provided by the *OED* is another reason it was chosen to be the data source for this study.

4. RESULTS

The results for this study are expansive, so this section will be split into an overview of the data collected for the entire time period, and several subsections focusing on each year selected for this study. When collecting the data, there were three types of discriminants I identified – lexical discriminants, quotation marks, and the use of italics – so the results focus on the frequency of these discriminants and how they change over time. For the purposes of this study, single and double quotation marks were included in the same category.

4. 1 GENERAL OVERVIEW

Looking at the data for all eight of the years selected for this study provides some insight into how discriminants have been used, and how they have changed, over a period of 350 years. To start, I will present the data for the overall frequency of new words signalled by discriminants over all 4000 entries analysed in this study.

Total Number of Word Entries Analysed	4000
Total Number of New Words Signalled by Discriminants	325
Frequency of New Words Signalled by Discriminants (%)	8.13%

Table 1. Overall frequency of new words signalled by discriminants in 4000 *OED* word entries

The frequency of new words signalled by discriminants is relatively low, at less than ten percent. However, this figure is still large enough to suggest that these discriminants are used with purpose, not just unconsciously, and it indicates that even in a smaller sample size, that discriminants could be productive in the retrieval of neologisms in texts. To better understand how the use of these discriminants has changed, it is important to look at the frequency of new words signalled by the author for each of the eight years I have chosen.

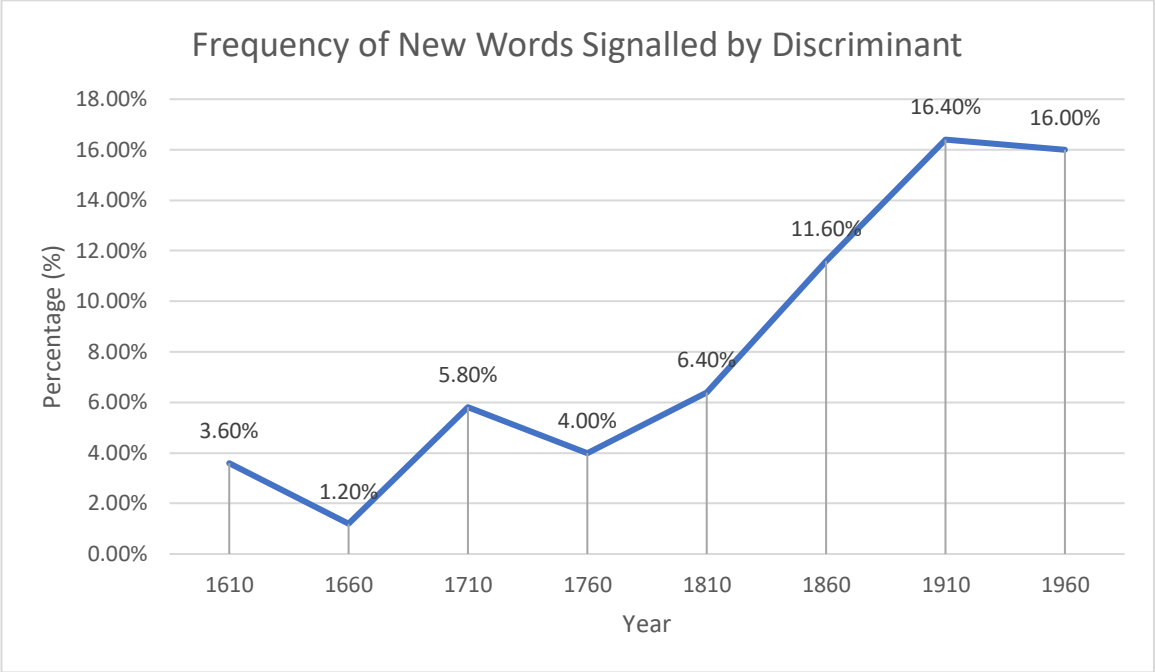

Figure 1. Graph showing the overall percentage of new words signalled by discriminants for each year

Year	Frequency of New Words Signalled by Discriminants (%)	Frequency of Words Signalled by Discriminants (per 1000 words)
1610	3.6%	36
1660	1.2%	12
1710	5.8%	58
1760	4.0%	40
1810	6.4%	64
1860	11.6%	116
1910	16.4%	164
1960	16.0%	160

Table 2. Table showing the overall percentage of new words signalled by discriminants for each year, and the average frequency of words signalled by discriminants per 1000 words

Figure **1** and Table **2** show the frequency of words signalled by discriminants for each of the eight years of the study. This shows that the frequency of new words being signalled by discriminants has increased by a large amount between 1610 and 1960, as I predicted. However, the data demonstrates that the number of new words signalled by discriminants does not increase with each year. Surprisingly, the frequency increases and decreases between the years of 1610 and 1760, with the values being quite close together.

This fluctuation was unexpected, as I had predicted that the years 1710 and 1760 would have the lowest overall frequency of discriminant usage due to the dip in new words being introduced to English during the eighteenth century.

The data demonstrates a steady increase in discriminant usage from 1760 onwards, with these values being notable enough to suggest that discriminants become more important from the nineteenth century. Impressively, for 1860, 1910, and 1960, over ten percent of the five hundred words analysed for each year were signalled to the audience using a discriminant. This suggests that authors found discriminants useful in introducing a new word from the nineteenth century onwards, perhaps as the overall number of new words being introduced to the language increased exponentially.

The results for 1910 and 1960 also show a levelling off, with a difference of only 0.4% between the two years, which could indicate that discriminant usage reaches its peak during this period, although this is not able to be guaranteed without any data from after 1960.

Ultimately these figures show an increase in discriminant usage over 350 years, but still demonstrates that the vast majority of new words are not signalled to an audience in any way.

The frequency of each discriminant type for each year is also an important piece of information, as it will provide insight into what the most common discriminant is, and how it has changed over time. This is shown in Figure **2**.

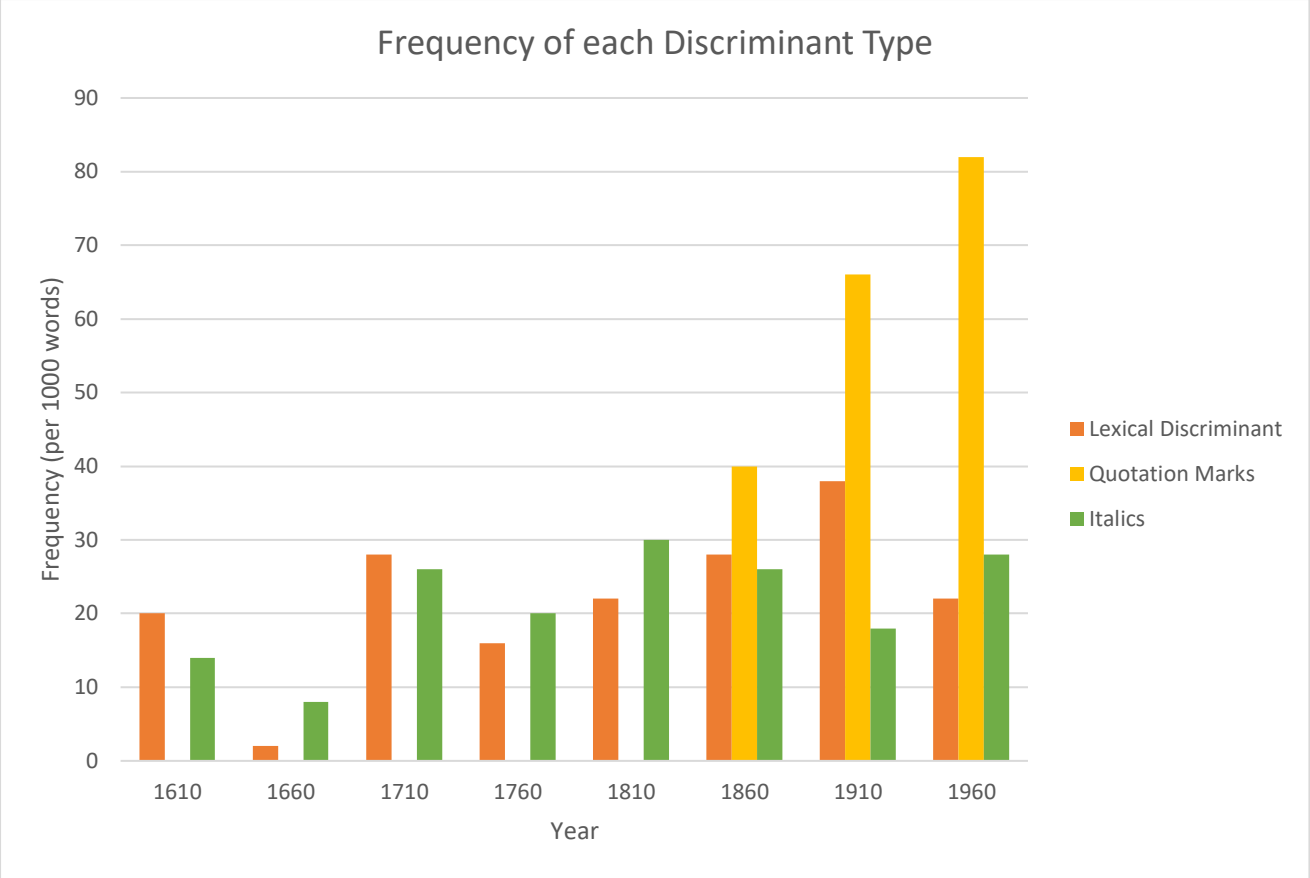

Figure 2. Graph demonstrating the frequency per 1000 words of each discriminant type for each year

	Frequency per 1000 words		
Year	**Lexical Discriminant**	**Quotation Marks**	**Italics**
1610	20	0	14
1660	2	0	8
1710	28	0	26
1760	16	0	20
1810	22	0	30
1860	28	40	26
1910	38	66	18
1960	22	82	28

Table 3. Table demonstrating the frequency per 1000 words of each discriminant type for each year

This data shows an interesting development in the use of the different types of discriminants over time. Both lexical discriminants and italics have been used to signal new words from 1610 up until 1960, but quotation marks were not found in the data I had collected until I reached the word entries for the year 1860. This is surprising, as I had predicted that quotation marks would be the most frequent discriminant type used throughout the entire time period I had selected. It is also a finding that is important in tracking how discriminants have been used over time, as this data shows that lexical discriminants and italics were initially favoured by authors.

The difference between lexical discriminants and the use of italics is admittedly marginal up until 1810, with the only large difference between the two types being displayed in the data for 1910. This could show that both discriminant types were viewed as equally helpful, as neither being largely favoured. There is also the possibility that these discriminants are equally signalling different types of new words, which would explain why they both have similar frequencies throughout. The types of new words that these discriminants signal will be explored later.

Despite not being present as a word signalling technique in the *OED* data until the year 1860, quotation marks immediately become the most frequent discriminant used. Importantly, the difference between the use of quotation marks and the use of lexical discriminants or italics is notable in 1860, and then far larger in 1910 and 1960. This would suggest that the three discriminant types aren't simply used interchangeably with no clear favoured technique, but that quotation marks become the most favoured way of signalling a new word to an audience if an author wants to signal the word. These findings for the later years correspond with the findings of Paryzek (2008), with quotation marks becoming the most productive way of signalling a new word, however the initial lack of quotation marks was not something I predicted. I believe that this finding is very important in tracing how discriminants have been used throughout the last few centuries in the English language, as it demonstrates a change in how new words were signalled, as well as seemingly demonstrating that eventually there became a collectively favoured discriminant.

When designing this study, I decided that analysing potential patterns between the types of new words introduced and the type of discriminant used to signal them would be very beneficial in understanding whether certain discriminants are used with a select purpose, or whether they are just used without much consideration as to what new word they are

signalling. The overview of the most frequent new word type found in the data for each type of discriminant is shown in Table **4**, only considering the words signalled by a single discriminant. The presence of multiple discriminants will be explored in the subsections discussing each year individually.

	Lexical Discriminant		**Quotation Marks**		**Italics**	
Year	**Most Frequent New Word Type**	**Percentage of Words Signalled by Discriminant**	**New Word Type**	**Percentage of Words Signalled by Discriminant**	**New Word Type**	**Percentage of Words Signalled by Discriminant**
1610	Borrowing	30.00%	-	-	Borrowing	57.14%
1660	Borrowing	100.00%	-	-	Borrowing	75.00%
1710	Compound	42.86%	-	-	Borrowing	53.85%
1760	Compound	75.00%	-	-	Borrowing	70.00%
1810	Compound	63.64%	-	-	Borrowing	53.33%
1860	Compound	35.71%	Compound	50.00%	Borrowing	61.54%
1910	Compound	63.16%	Compound	48.48%	Borrowing	66.67%
1960	Compound	54.55%	Compound	46.34%	Borrowing	57.14%

Table 4. Table showing the most frequent new word type signalled by each discriminant type, per year

This table shows only two new word types, indicating that the most likely new word types to be signalled by a discriminant are compounds and borrowings. This is unsurprising when considering words borrowed from other languages, as they could be accompanied by explanations or definitions, meaning that specifically signalling the word would be beneficial to whoever is reading the text. In relation to other new word types, borrowings can be the least transparent, but compounds can also lack transparency, as I have already discussed. Therefore, this data could indicate that discriminants could be used by authors more when the word they are introducing are less transparent in their meaning. However, borrowings and compounds are also amongst the most common types of new words in the English language,

meaning that this spread of data could be a result of these word types making up a bigger percentage of the 500 entries analysed per year.

The most frequent word type signalled by lexical discriminants changes is shown to change from 1660 to 1710, going from borrowings to compounds. As the frequency of borrowings signalled by lexical discriminants in relation to other word types signalled by lexical discriminants changes rapidly between 1610 and 1660, it is likely that this is a product of relatively few words being signalled by discriminants at all during this period, therefore I would not take this to mean that lexical discriminants once appeared alongside borrowings more frequently, and then there was a collective intentional shift towards compounds.

The fact that compounds remain the most frequent word type alongside lexical discriminants, however, does appear to be important, especially when considering the years beyond 1810, as this is when more discriminants were being used overall. Despite being a very common new word formation technique, the fact that the frequency of compounds and lexical discriminants is over half of all instances of lexical discriminants being used in 4 separate years suggests that there is a greater tendency for lexical discriminants to be used when signalling a compound. As this study only considered a limited sample size, this could be a very interesting avenue for further research.

Compounds are also shown to be signalled by quotation marks consistently from 1810 onwards, with frequencies of half, or close to half, of all new words signalled by quotation marks. Again, whilst taking into consideration the overall frequency of compounds in relation to other word types, this data still suggests that authors could favour the use of quotation marks when signalling compounds. When considering the trends of both lexical discriminants and quotation marks, it appears that, in general, both types of discriminants could be favoured when introducing a new compound, and this could be very helpful knowledge for the future of neologism retrieval research.

Finally, this table also shows that borrowings are the most frequent word type signalled by italics throughout the 350-year time period. Interestingly, the frequency of borrowings being signalled by italics remains above fifty percent of all word types signalled by italics, for all of the years analysed in this study. This would suggest that the use of italics is favoured by authors when introducing a borrowing, perhaps because the text style helps indicate that the word is in a different language compared to the other words accompanying it in the text. The consistent correlation between borrowings and italics indicates that there is a level of understanding amongst authors of what certain discriminants express, and it could be argued that this data shows that authors understand that borrowings are best expressed

through the use of italics. This is a correlation that could also be very important to further discriminant and neologism research.

Another key focus of this study was the range of lexical discriminants that could be used to signal a new word, in order to expand on the few that have been mentioned in previous research. The results for the total unique discriminants per year can be seen in Figure **3**.

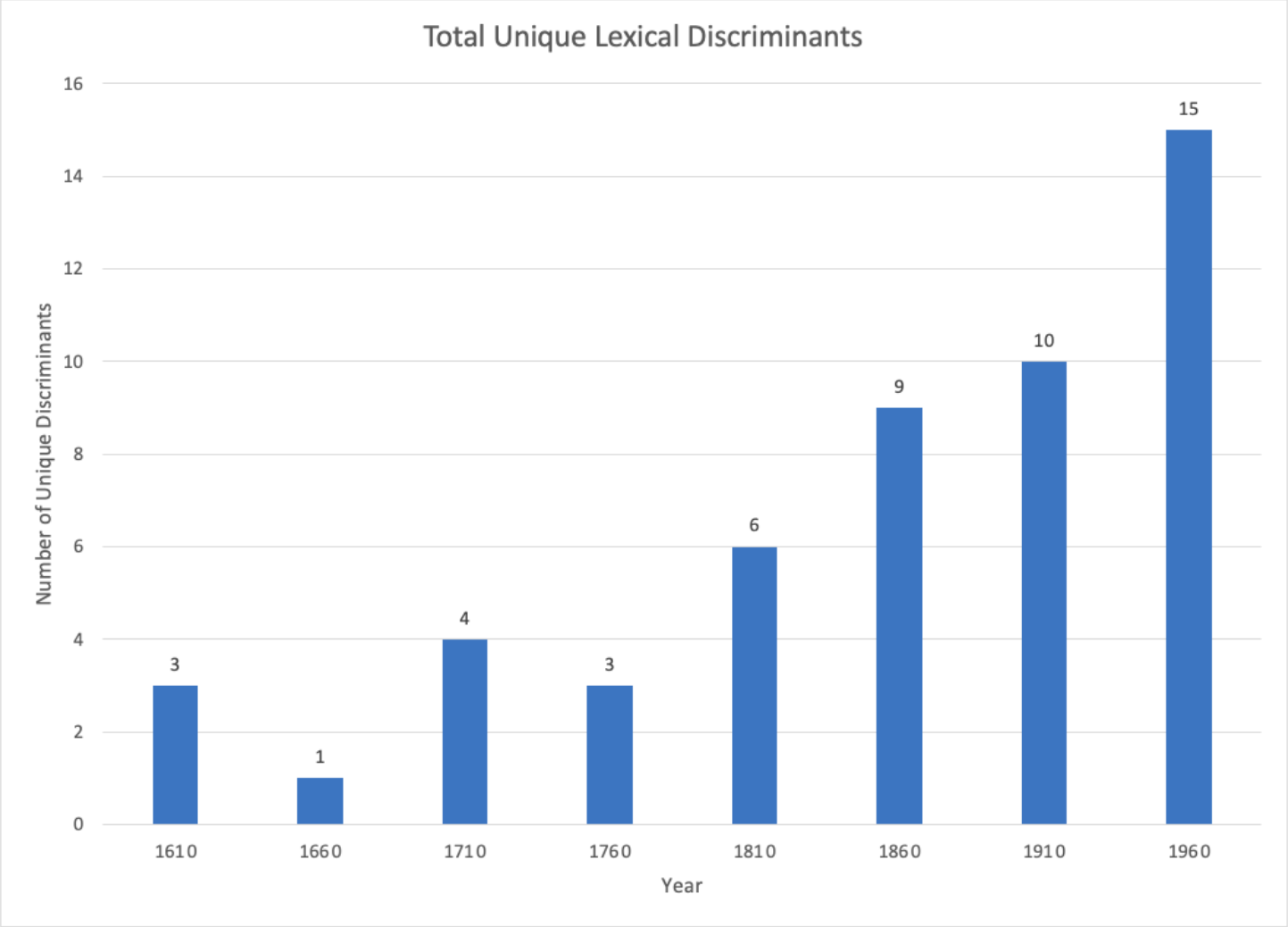

Figure 3. Graph showing the total unique instances of lexical discriminants per year

This demonstrates the same initial fluctuation that was seen in the overall frequency of discriminants in Figures **1** and **2**, with there being no steady increase until 1810. As there was a lower number of lexical discriminants being used during these earlier years anyways, it is not surprising that the range of lexical discriminants used remains small. Even though there

is only a small number of different words and phrases used to signal a new word in texts during these years, there is only one year in which there is only one type of lexical discriminant found. This shows that even when there are few examples of lexical discriminants in texts, there are still multiple words or phrases that can be used to signal new words, proving that the range of lexical discriminants used in the field of neologism retrieval and research could be expanded.

It also follows that the range of lexical discriminants used alongside new words has expanded as the rate of new words being added to the English language increase, indicating that authors have found new ways to draw their audiences' attention to certain words over the years. The steadily growing number of lexical discriminants in use after 1760 therefore appears to reflect the steadily growing English language, and it could also reflect that as society has expanded and evolved, so to have the ways in which authors communicate with their audience.

Year	**Most Frequent Lexical Discriminant**	**Frequency of Total Lexical Discriminants (%)**	**Frequency of Lexical Discriminant (per 1000 words)**
1610	called	72.00%	16
1660	called	100.00%	4
1710	called + call(s)	43.75% + 43.75%	14 + 14
1760	called	70.00%	14
1810	called	50.00%	16
1860	called	45.83%	22
1910	called	30.00%	24
1960	called	48.00%	6

Table 5. Table showing the most frequent lexical discriminant for each year

Table **5** shows the most frequent lexical discriminant for each year, and here we can see that the discriminant *called* remains the most frequent lexical discriminant throughout the selected time period. This follows my prediction, and shows agreement with the findings of Paryzek, that *called* is one of the most productive lexical discriminants. Therefore, this

suggests that when searching for neologisms, using this discriminant would be very beneficial in this search.

The prevalence of *called* is also interesting as it demonstrates that a range of sources all use similar words and phrases to introduce new words, despite the fact that discriminants appear to be used as a steppingstone with very little attention paid to them beyond their use in automated neologism retrieval. This trend could suggest that there is a wide understanding of how discriminants work and how they can be best utilised amongst anyone who is introducing a new word in a text, even if they seem to be a technique that is very rarely discussed. There is a tentativeness required when exploring the sentiment of an almost universal understanding of discriminants and their usage by authors throughout time, mainly due to the large number of new words that were analysed as part of this study that were not signalled by any discriminant. So perhaps it is safer to suggest that this evident favouring of the discriminant *called* is a greater indication that it is accepted as the easiest, and potentially best, way of introducing a new item or concept to an audience that they believe has not encountered the word before.

Now that I have given the outline of the data across the entire 350-year period, analysing the data I collected for each year may provide greater insight into the range, or lack of, in how discriminants are used and what type of words they are used with, along with a list of the different lexical discriminants found within each set of the 500 word entries, and some observations on how multiple discriminants may be used together to signal one word.

4.2.1. DISCRIMINANT USAGE – 1610

4.2.1.1 DISCRIMINANT FREQUENCY

As seen in Figure **2** and Table **3,** the only discriminants found in the data for 1610 were lexical discriminants and italics, with lexical discriminants being the most productive of the two. Table **6** shows the frequency of each lexical type in the data for the year 1610, including the one example of two discriminants being used together to signal one word.

Discriminant Type	Occurrences	Frequency of Occurrences in 500 Entries (%)	Frequency (per 1000 words)
Lexical Discriminant	10	2%	20
Quotation Marks	0	0	0
Italics	7	1.4%	14
Lexical Discriminant + Italics	1	0.2%	2

Table 6. Frequency of all discriminants out of 500 *OED* entries for the year 1610

Again, it is clear that there is not a large difference between the usage of lexical discriminants and italics at this stage. The one instance of a lexical discriminant being used alongside displaying the new word in italics in the text, however, seems to show that there is some level of intent behind the usage of discriminants, even in this early time period. Using two discriminants together would imply that the author of this text is aware of the multiple discriminants that can be used in signalling potentially unknown words, and that they are potentially using more than one type of discriminant in order to really communicate how different this new word is. I believe that this is important to consider, as it not only provides some insight into how discriminants are used in the text itself, but also allows for the start of some investigation into how these techniques are used as a key component of communication between people.

4.2.1.2 LEXICAL DISCRIMINANTS

There are only three different lexical discriminants present within the ten instances of lexical discriminant signalling for this year, with the most frequent being *called*. There is a notable difference in frequency between *called* and the second most frequent lexical discriminant *termed*, which would suggest that *called* is favoured when an author decides a new word needs to be signalled through a lexical discriminant, but the overall number of signalling for this year is one of the smallest.

Lexical Discriminant	Occurrences	Percentage of Occurrences in 500 Entries	Frequency per 1000 words
called	8	1.6%	16
termed	2	0.4%	4
call(s)	1	0.2%	2

Table 7. All lexical discriminants found in the *OED* sample for the year 1610

4.2.1.3 WORD TYPES

Now focusing on the different word types signalled by each discriminant can show widely they are used, or whether they are only used for one or two different word types.

Discriminant Type: Lexical			
Word Formation Type	**Occurrences**	**Percentage of Occurrences in 500 Entries**	**Frequency per 1000 words**
Borrowing	3	0.6%	6
Compound	2	0.4%	4
Proper Noun	2	0.4%	4
Prefixation	1	0.2%	2
Suffixation	1	0.2%	2
Variant	1	0.2%	2
Uncertain	1	0.2%	2

Table 8. Each word type signalled by only lexical discriminants for the year 1610

Table **8** shows that lexical discriminants are used for a range of word types, with no large difference in the frequencies that would suggest at this point that lexical discriminants are purposefully used more frequently alongside one specific word type. I think that this spread of data demonstrates that lexical discriminants can be very versatile in what they are signalling to an audience, which would imply that using these discriminants for automatic neologism retrieval could lead to the finding of varied types of neologisms.

Discriminant Type: Italics			
Word Formation Type	**Occurrences**	**Percentage of Occurrences in 500 Entries**	**Frequency per 1000 words**
Borrowing	4	0.8%	8
Unknown	2	0.4%	4
Suffixation	1	0.2%	2

Table 9. Each word type signalled by only italics for the year 1610

Discriminant Type: Lexical Discriminant + Italics			
Word Formation Type	**Occurrences**	**Percentage of Occurrences in 500 Entries**	**Frequency per 1000 words**
Suffixation	1	0.2%	2

Table 10. Each word type signalled by a combination of lexical discriminants and italics for the year 1610

Tables **9** and **10** show that there is less variation of word types signalled by italics, but again this is due to the overall lack of discriminants during the earlier years. As discussed previously, borrowings are the most common word type signalled by italics for this year. The most interesting aspect is the notion of lexical discriminants and italics being used together, although not much can be said for this in relation to the link between word type and discriminant type as there was only one occurrence of multiple discriminants in all 500 *OED* entries analysed for 1610.

4.2.2 DISCRIMINANT USAGE – 1660

4.2.2.1 DISCRIMINANT FREQUENCY

Due to the surprising dip in discriminant usage in 1660 compared to 1610, there is very little to analyse in depth for this year. Interestingly, there is still one occurrence of lexical discriminants and italics being used together to signal one new word, which could show that the combination of the two could have potential to become an established way of signalling new words.

Discriminant Type	Occurrences	Percentage of Occurrences in 500 Entries	Frequency per 1000 words
Lexical Discriminant	1	0.2%	2
Quotation Marks	0	0	0
Italics	4	0.8%	8
Lexical Discriminant + Italics	1	0.2%	2

Table 11. Frequency of all discriminants in 500 *OED* entries for the year 1660

4.2.2.2. LEXICAL DISCRIMINANTS

The only lexical discriminant found in the entries for 1660 was *called*, with 2 total occurrences. This can prove that *called* could be the lexical discriminant favoured for new word signalling over a large period of time, but it is not a permanent conclusion that can be made on a reasonably small data set.

Lexical Discriminant	Occurrences	Percentage of Occurrences in 500 Entries	Frequency per 1000 words
called	2	0.4%	4

Table 12. All lexical discriminants found in the *OED* sample for the year 1660

4.2.2.3 WORD TYPES

Much like the other sections of the data for the year 1660, very little can be said any potential link between word types and discriminants, as there were only a few instances of signalling found. This does open up the opportunity for further research using a greater number of sources, alongside the possibility of considering the sociohistorical context for this time period but this is not very helpful for this particular study.

Discriminant Type: Lexical			
Word Formation Type	**Occurrences**	**Percentage of Occurrences in 500 Entries**	**Frequency per 1000 words**
Borrowing	1	0.2%	2

Table 13. Each word signalled by only lexical discriminants in the *OED* sample for the year 1660

Discriminant Type: Italics			
Word Formation Type	**Occurrences**	**Percentage of Occurrences in 500 Entries**	**Frequency per 1000 words**
Borrowing	3	0.6%	6
Compound	1	0.2%	2

Table 14. Each word signalled by only italics in the *OED* sample for the year 1660

Discriminant Type: Lexical Discriminant + Italics			
Word Formation Type	**Occurrences**	**Percentage of Occurrences in 500 Entries**	**Frequency per 1000 words**
Unknown	1	0.2%	2

Table 15. Each word signalled by a combination of lexical discriminants and italics in the *OED* sample for the year 1660

4.2.3 DISCRIMINANT USAGE – 1710

4.2.3.1 DISCRIMINANT FREQUENCY

The data in Table **16** shows that lexical discriminants and italics were very close in the year 1710, and that there is still a combined use of lexical discriminants and italics being used to signal a single word. At this point it appears that lexical discriminants and italics have equal favour, although the analysis of word types in the upcoming section will show that there is a difference in the variety of words these discriminants are being used to signal.

Discriminant Type	Occurrences	Percentage of Occurrences in 500 Entries	Frequency per 1000 words
Lexical Discriminant	14	2.8%	28
Quotation Marks	0	0	0
Italics	13	2.6%	26
Lexical Discriminant + Italics	2	0.4%	4

Table 16. Frequency of all discriminants in 500 *OED* entries for the year 1710

4.2.3.2 LEXICAL DISCRIMINANTS

There is one new lexical discriminant used in this sample that has not been used in the previous years, which is the phrase *this is a new word*. This is rather more direct than the other lexical discriminants found in this year – *called, call(s),* and *termed* – which I believe shows that there are varying degrees of severity to lexical discriminants, and that authors appear to prefer the slightly more subtle ways of communicating a new word to an audience. It could also demonstrate that lexical discriminants are used to convey both brand new words that have not been used elsewhere, and very rarely used words that have only been recently introduced to the language that the author is guessing will be unfamiliar to a vast majority of their audience.

Lexical Discriminant	Occurrences	Percentage of Occurrences in 500 Entries	Frequency per 1000 words
called	7	1.4%	14
call(s)	7	1.4%	14
this is a new word	1	0.2%	2
termed	1	0.2%	2

Table 17. All lexical discriminants found in the *OED* sample for the year 1710

4.2.3.3 WORD TYPES

The data for this year shows a higher frequency of compounds and borrowings being signalled by only lexical discriminants compared to other word types, with some instances of

words formed through prefixation, suffixation, and conversion also being signalled with this technique.

Discriminant Type: Lexical			
Word Formation Type	**Occurrences**	**Percentage of Occurrences in 500 Entries**	**Frequency per 1000 words**
Compound	6	1.2%	12
Borrowing	5	1.0%	10
Prefixation	1	0.2%	2
Suffixation	1	0.2%	2
Conversion	1	0.2%	2

Table 18. Each word signalled only by lexical discriminants in the *OED* sample for the year 1710

The word types signalled by only italics also show some variation, although it is clearly more frequently used alongside borrowings than any other word type. Again there are examples of lexical discriminants being used alongside italics, showing that multiple discriminants can be used together for the same purpose, but this is still a tiny percentage of the total number of signalling techniques found in this year.

Discriminant Type: Italics			
Word Formation Type	**Occurrences**	**Percentage of Occurrences in 500 Entries**	**Frequency per 1000 words**
Borrowing	7	1.4%	14
Uncertain	2	0.4%	4
Prefixation	1	0.2%	2
Compound	1	0.2%	2
Variant	1	0.2%	2

Loan Blend	1	0.2%	2

Table 19. Each word signalled only by italics in the *OED* sample for the year 1710

Discriminant Type: Lexical Discriminant + Italics			
Word Formation Type	**Occurrences**	**Percentage of Occurrences in 500 Entries**	**Frequency per 1000 words**
Borrowing	1	0.2%	2
Uncertain	1	0.2%	2

Table 20. Each word signalled by a combination of lexical discriminants and italics for the year 1710

4.2.4 DISCRIMINANT USAGE – 1760

4.2.4.1 DISCRIMINANT FREQUENCY

The discriminant frequency for the 1760 sample is not very different from the years that came before, showing some consistency in usage, despite the relatively small frequencies, including the use of lexical discriminants combined with italics.

Discriminant Type	**Occurrences**	**Percentage of Occurrences in 500 Entries**	**Frequency per 1000 words**
Lexical Discriminant	8	1.6%	16
Quotation Marks	0	0	0
Italics	10	2.0%	20
Lexical Discriminant + Italics	2	0.4%	4

Table 21. Frequency of all discriminants in the *OED* sample for the year 1760

4.2.4.2 LEXICAL DISCRIMINANTS

There is very little range in the number of unique discriminants used in the 1760 sample, with only three being present. Unsurprisingly, the less frequent *call(s)* and *known by the name of* are less frequent to a notable degree, leaving no question as to whether the discriminant *called* is intentionally favoured when signalling a new word.

Lexical Discriminant	**Occurrences**	**Percentage of Occurrences in 500 Entries**	**Frequency per 1000 words**
called	7	1.4%	14
call(s)	2	0.4%	4
known by the name of…	1	0.2%	2

Table 22. All lexical discriminants found in the *OED* sample for the year 1760

4.2.4.3 WORD TYPES

As with the previous years, the most prevalent word types signalled by discriminants are borrowings and compounds, far outweighing the presence of other word types that were also signalled. Interestingly, this year demonstrates a clearer favouring of lexical discriminants when a compound is signalled, instead of the usual mix of compounds and borrowing. The sample data for this year clearly shows that italics signal borrowings first and foremost, but that doesn't prevent them from being used to signal other word types.

Discriminant Type: Lexical			
Word Formation Type	**Occurrences**	**Percentage of Occurrences in 500 Entries**	**Frequency per 1000 words**
Compound	6	1.2%	12
Borrowing	1	0.2%	2
Suffixation	1	0.2%	2
Discriminant Type: Italics			
Word Formation Type	**Occurrences**	**Percentage of Occurrences in 500 Entries**	**Frequency per 1000 words**

Borrowing	7	1.4%	14
Suffixation	1	0.2%	2
Loan Blend	1	0.2%	2
Proper Name and Suffix	1	0.2%	2
Discriminant Type: Lexical + Italics			
Word Formation Type	**Occurrences**	**Percentage of Occurrences in 500 Entries**	**Frequency per 1000 words**
Borrowing	2	0.4%	4

Table 23. All word types signalled by discriminants in the *OED* sample for the year 1760

4.2.5 DISCRIMINANT USAGE – 1810

4.2.5.1 DISCRIMINANT FREQUENCY

The overall discriminant frequency in 1810 is slightly larger than anything that has come before it but shows much of the same patterns that have already been observed elsewhere. There is the first occurrence of quotation marks used to signal new words, but this is used in combination with a lexical discriminant. Again, the difference in frequency between lexical discriminants and italics is negligible.

Discriminant Type	**Occurrences**	**Percentage of Occurrences in 500 Entries**	**Frequency per 1000 words**
Lexical Discriminant	11	2.2%	22
Quotation Marks	0	0	0
Italics	15	3.0%	30
Lexical Discriminant + Italics	4	0.8%	8
Lexical Discriminant + Quotation Marks	1	0.2%	2

Table 24. Discriminant frequency in the *OED* sample for the year 1810

4.2.5.2 LEXICAL DISCRIMINANTS

This year at the start of the nineteenth century demonstrates a greater range of unique discriminants than any year that has come before it in this study, with *called* and *call(s)* still being the favoured options. Interestingly, there is the return of the discriminant *known by the name* which only appeared once in the sample for 1760, as well as the appearance of two new discriminants *gave it the name* and *given it the name*. This indicates a potential growing awareness of the range of lexical discriminants that could signal a new word.

Lexical Discriminant	**Occurrences**	**Percentage of Occurrences in 500 Entries**	**Frequency per 1000 words**
called	8	1.6%	16
call(s)	4	0.8%	8
termed	1	0.2%	2
known by the name…	1	0.2%	2
gave it the name	1	0.2%	2
given it the name	1	0.2%	2

Table 25. Lexical discriminants found in the *OED* sample for the year 1810

4.2.5.3 WORD TYPE

There is no notable change in the trend already seen in the previous years, with both compounds and borrowings being the most signalled word types. Again, compounds are most frequently signalled with lexical discriminants, and borrowings are most frequently signalled with italics. The combination of lexical discriminants and quotation marks is also used to signal a borrowing, although not much can be discussed in relation to this as it only occurs once.

4.2.6 DISCRIMINANT USAGE – 1860

4.2.6.1 DISCRIMINANT FREQUENCY

The year 1860 shows a notable increase in overall discriminant frequency compared to the previous years, and also shows the first instances of quotation marks being used without the presence of other types of discriminants. It is also interesting to note that at some point between 1810 and 1860, quotation marks appear to have become the most favoured way of signalling a new word in a text to an audience, despite only being used once beforehand. This seems to indicate some kind of shift in texts between these two years, in which quotation marks became an accepted and understood method of presenting a new word.

Discriminant Type	**Occurrences**	**Percentage of Occurrences in 500 Entries**	**Frequency per 1000 words**
Lexical Discriminant	14	2.8%	28
Quotation Marks	20	4.0%	40
Italics	13	2.6%	26
Lexical Discriminant + Italics	3	0.6%	6
Lexical Discriminant + Quotation Marks	7	1.4%	14
Italics + Quotation Marks	1	0.2%	2

Table 26. Discriminant frequency found in the *OED* sample for the year 1860

4.2.6.2 LEXICAL DISCRIMINANTS

Table **27** shows that there is a continued increase in the number of unique lexical discriminants used within the sample quotes of this study, but that the frequency of these lexical discriminants is still very low compared to the frequency of *called*. The increasing variation with every year is promising, showing just how many different lexical discriminants can be used to signal new words.

Lexical Discriminant	Occurrences	Percentage of Occurrences in 500 Entries	Frequency per 1000 words
called	11	2.2%	22
the name of…	4	0.8%	8
call(s)	1	0.2%	2
so-called	1	0.2%	2
propose the word	1	0.2%	2
proposes to name it	1	0.2%	2
I name…	1	0.2%	2
termed	1	0.2%	2
the name	1	0.2%	2

Table 27. All lexical discriminants found in the *OED* sample for the year 1860

4.2.6.3 WORD TYPES

The data for word types in 1860 maintains the same consistency as seen in previous years with borrowings and compounds being the most frequent types signalled by discriminants. Again, borrowings are the word type most frequently signalled by italics, and compounds are the most frequent elsewhere.

Compounds are also notably more frequently signalled by quotation marks than any other word type, once again indicating that authors, for whatever reason, want to bring their audiences' attention to new compounds more than almost any other word type. It is clear that by this stage in the study, the trends I have already identified will continue into the years 1910 and 1960, as the data compiled for each of the eight years very simply reflects the overall averages discussed in the first part of this results section.

Discriminant Type: Quotation Marks			
Word Formation Type	**Occurrences**	**Percentage of Occurrences in 500 Entries**	**Frequency per 1000 words**
Compound	10	2.0%	20
Suffixation	3	0.6%	6
Borrowing	3	0.6%	6
Conversion	2	0.4%	4
Prefixation	1	0.2%	2
Loan Blend	1	0.2%	2

Table 28. All word types signalled by quotation marks in the *OED* sample for 1860

As the overall instances of discriminants increases, so does the range of word types signalled by the discriminants, demonstrating that compounds are simply the favoured word type, and that discriminants are not wholly exclusive to any one type of word.

4.2.7 DISCRIMINANT USAGE – 1910

4.2.7.1 DISCRIMINANT FREQUENCY

The largest increase in overall discriminant frequency so far is found in the data for 1910, with a larger frequency of quotation marks and lexical discriminants, compared to italics. This could indicate that as the use of quotation marks has grown more frequent, and they have become more accepted as a discriminant, that there are less instances in which italics feel like the best option.

There are also more occurrences of multiple discriminants being used at one time, enough so that it would appear this combination of discriminants has become an important part of new word signalling in itself and is indeed used intentionally in order to highlight specific words in a more effective way than a single discriminant could achieve.

Discriminant Type	Occurrences	Percentage of Occurrences in 500 Entries	Frequency per 1000 words
Lexical Discriminant	19	3.8%	38
Quotation Marks	33	6.6%	66
Italics	9	1.8%	18
Lexical Discriminant + Italics	9	1.8%	18
Lexical Discriminant + Quotation Marks	12	2.4%	24

Table 29. Discriminant frequency found in the *OED* sample for the year 1910

4.2.7.2 LEXICAL DISCRIMINANTS

Continuing the upwards trend of the previous years, the data for 1910 shows another increase in the range of unique lexical discriminants used to signal a new word. This time there are ten unique lexical discriminants, with *called* still being the most frequent by far.

Over half of the ten lexical discriminants found in this sample contain the word *name*, presenting another lexical item that could yield results in an automatic or manual neologism search in a corpus, even if it is not the most productive of the lexical discriminants.

The discriminants in Table **30** also correspond with slightly different neologism concepts. The discriminants *known as* appear to signal some words that have been used before, but not to any significant extent as they still remain to a wider audience, whilst the discriminants *designate as* and *I name* are used alongside words that the author of the text has coined themselves. Therefore, this shows that discriminants can very subtly communicate with the audience exactly what kind of neologism they are being presented with.

Lexical Discriminant	Occurrences	Percentage of Occurrences in 500 Entries	Frequency per 1000 words
called	12	2.4%	24
known as	8	1.6%	16
call(s)	3	0.6%	6
designate as	2	0.4%	4
have been named	1	0.2%	2
propose(s) the name	1	0.2%	2
receive the name of	1	0.2%	2
classify them under the collective name	1	0.2%	2
I name	1	0.2%	2
name… proposed	1	0.2%	2

Table 30. All lexical discriminants found in the *OED* sample for the year 1910

4.2.7.3 WORD TYPES

The most frequent word types signalled by discriminants remains compounds and borrowings, with compounds being the most frequent word type signalled by both lexical discriminants and quotation marks. Surprisingly there is an increase in the frequency of new words formed by suffixation being signalled by quotation marks compared to previous years. This again serves as a reminder that, for all these discriminants are used most frequently with two new word types, there is no exclusivity in any of the discriminants. They remain versatile and are able to be used with any new word, which is why they are such a fascinating concept.

Borrowings remain the most frequent word type used alongside italics, and loan blends are seen to be most frequently signalled by a combination of lexical discriminants and italics. As loan blends are a combination of a borrowed word and some English elements, it makes sense as to why multiple discriminants would be utilised in order to highlight the neologism to an audience.

Discriminant Type: Quotation Marks			
Word Formation Type	**Occurrences**	**Percentage of Occurrences in 500 Entries**	**Frequency per 1000 words**
Compound	16	3.2%	32
Suffixation	7	1.4%	14
Borrowing	5	1.0%	10
Clipping	3	0.6%	6
Conversion	2	0.4%	4

Table 31. All word types signalled by quotation marks in the *OED* sample for the year 1910

Discriminant Type: Lexical + Italics			
Word Formation Type	**Occurrences**	**Percentage of Occurrences in 500 Entries**	**Frequency per 1000 words**
Loan Blend	3	0.6%	6
Compound	2	0.4%	4
Proper Name and Suffix	2	0.4%	4
Borrowing	1	0.2%	2
Unknown	1	0.2%	2

Table 32. All word types signalled by a combination of lexical discriminant and italics in the *OED* sample for the year 1910

4.2.8 DISCRIMINANT USAGE – 1960

4.2.8.1 DISCRIMINANT FREQUENCY

The data for the final year analysed in this study appears to demonstrate the start of a stabilisation in the use of discriminants to signal new words, with the frequencies being very similar to those seen in the year 1910. Quotation marks remain the most frequent discriminant type, with italics, lexical discriminants, and the combination of lexical discriminants and quotation marks.

Discriminant Type	Occurrences	Percentage of Occurrences in 500 Entries	Frequency per 1000 words
Lexical Discriminant	11	2.2%	22
Quotation Marks	41	8.2%	82
Italics	14	2.8%	28
Lexical Discriminant + Italics	2	0.4%	4
Lexical Discriminant + Quotation Marks	12	2.4%	24

Table 33. Discriminant frequency found in the *OED* sample for the year 1960

4.2.8.2 LEXICAL DISCRIMINANTS

Here we see the largest number of unique lexical discriminants, with fifteen found in the sample for 1960. The most frequently used remains *called.* Once again, the discriminants in this sample can signal either new words that have been used only a few times that an author is presuming their audience has never encountered before (i.e. *called, termed, referred to as*), or words that the author has coined themselves, as exhibited in the use of *coin the term* and *propose the name*.

Lexical Discriminant	Occurrences	Percentage of Occurrences in 500 Entries	Frequency per 1000 words
called	6	1.2%	12
termed	3	0.6%	6
known as	2	0.4%	4
so-called	2	0.4%	4
referred to as	2	0.4%	4
capsuled as	1	0.2%	2
coin the term	1	0.2%	2
coined word	1	0.2%	2
denoted	1	0.2%	2
labelled	1	0.2%	2
named	1	0.2%	2
new drug	1	0.2%	2
new term	1	0.2%	2
propose the name	1	0.2%	2
suggest the term	1	0.2%	2

Table 34. All lexical discriminants found in the *OED* sample for the year 1960

4.2.8.3 WORD TYPES

There is little change between 1910 and 1960 in regard to the most frequent word type signalled by the discriminants, with compounds still being the most frequent word type signalled by lexical discriminants, quotation marks, and a combination of the two. Borrowings also remain the most frequent word type signalled by italics.

What has changed is the range of word types signalled by quotation marks. Although many of these word types have a comparatively low frequency, it still shows that these discriminants can signal any word type, no matter how different or complex they may be. That this range is demonstrated the most at this point in time could indicate that discriminants have come to be used more widely as new items and concepts are created rapidly in a society where access to information is expanding, but this is not a conclusion that can be reached with any certainty based on only one year of data.

Discriminant Type: Quotation Marks			
Word Formation Type	**Occurrences**	**Percentage of Occurrences in 500 Entries**	**Frequency per 1000 words**
Compound	19	3.8%	38
Suffixation	6	1.2%	12
Blend	5	1.0%	10
Prefixation	3	0.6%	6
Borrowing	3	0.6%	6
Clipping	2	0.4%	4
Loan Blend	1	0.2%	2
Acronym	1	0.2%	2
Variant	1	0.2%	2

Table 35. All word types signalled by quotation marks in the *OED* sample for the year 1960

4.3 OVERALL RESULTS

Overall, it is clear that discriminants have been used to signal new words since at least 1610, with the frequency of their use steadily increasing from 1810 onwards. The data has shown that lexical discriminants and italics were used before quotation marks, with them having a similar frequency to one another, for almost all of the years analysed in this study. By 1860 quotation marks became the most frequent discriminant for signalling new words, far outpacing the use of lexical discriminants and italics. The popularity of the quotation marks as a signalling device was something I had predicted, in line with the research of Paryzek (2008), but I had not predicted that there would be a large period of time where quotation marks were not present in this capacity at all.

It has also been found that there are instances of multiple discriminants being used together to signal the same word in each year of this study, and the frequency of these occurrences increased over time, much like the frequency of the singular use of discriminants. Lexical discriminants are shown to be used with either italics or quotation marks, and the data indicates that the combination of italics and quotation marks to signal one word is avoided.

As I predicted, this study found many lexical discriminants that hadn't been identified in previous research. The number of lexical discriminants increased over time, and the discriminant *called* was the most frequent lexical discriminant for each year of the study, keeping in line with my prediction. However, I had also predicted that the discriminant *termed* would be the second-most productive lexical discriminant in this study and that was not the case, although it was used in multiple years.

It has also been shown that borrowings and compounds are the most frequent new word types signalled by discriminants, with compounds becoming the most frequent word type signalled by lexical discriminants and quotation marks, and borrowings remaining the most frequent word type signalled by italics over the 350-year period that was the focus of this study. Due to the scope of this study, it remains unclear as to whether lexical discriminants and quotation marks are used to signal compounds to a greater extent than any other word type for any reason other than the fact that compounding is a highly productive word formation process and therefore a large quantity of new words are compounds.

5 DISCUSSION

This study successfully tracked the use of multiple types of discriminants across the 350-year period selected, identifying various different lexical discriminants, and observing how discriminant usage has expanded.

The results of this study into discriminants has started to fill in some of the gaps in the research regarding discriminants, and has afforded them some more focus outside of only their use as neologism retrieval tools, although their productivity in this area has not been, and cannot be, ignored. The lexical discriminants found in this study could be used to expand the search functions of automatic neologism retrieval programmes, and I believe that the general overview of the discriminant use over 350 years is a useful indication as to which years have been important for neologism introduction and signalling and provides several years to which more focus and in-depth research could be afforded. It has been shown that discriminants are extremely useful in identifying neologisms manually, too, especially in the case of the use of italics which may not be able to be utilised in automated retrieval software. I believe that discriminants are best researched in a manual method, as to get a better understanding of their range instead of just searching for the frequency of a few select discriminants, which is potentially why so little attention has been paid to them in previous research.

Researching discriminants themselves means that a large amount of research in this area will be manual, which can be time consuming and leads to smaller sample sizes than could be achieved through automated methods. Thankfully, the *OED* is very easy to access and provides the relevant information in such a way that analysing the 500 word entries for each year was not complex. If a larger research project was undertaken, I believe that the *OED* would still be a very useful tool, and other dictionaries could also be considered in order to piece together a picture of how discriminants have been used. This study has already touched upon how the smaller sample size used allowed for the observation of trends, mainly the frequency of compounds and borrowings throughout, but did not allow for conclusions of causation to be made. I believe that a study that analyses the word types of all words in a sample, and not just those that have been signalled by a discriminant, could help to determine whether the prevalence of compounds being signalled by lexical discriminants and quotation marks in this study is a result of the increased frequency of new words formed by compounding in general, or whether compounds are intentionally signalled through the use of these discriminants.

Any research into discriminants using *OED* data is ultimately at the mercy of whatever has been selected by lexicographers to appear in a dictionary, which is why this study acknowledged that the discriminants found in this study were either signalling words created by the author of the text themselves, or very rarely used words that the author is aware their audience will likely not know. Whilst very useful, unfortunately the *OED* first attestation quotes do not account for new words that had been used solely verbally before they were put to text. Further research into discriminant usage in modern-day texts may also be difficult using dictionaries due to the lack of present-day neologisms being added as they are being used, meaning that a focus on the historical use of discriminants could be more beneficial at this time.

Finally, I believe that future research into discriminant usage has many avenues aside from the ones I have already named. As well as word types, I believe that analysing how the text type a new word appears in could influence the type of discriminant used could help form a greater understanding of how these techniques work, as well as analysing how the types of discriminants found in certain text types has developed over a similar time period as to the one used in this study.

6 CONCLUSION

In conclusion, this study has aimed to delve into how a range of discriminants have been used alongside neologisms over a 350-year period, hoping to shed some light on their frequency and how they have developed over time. I used the small research piece conducted by Paryzek (2008) as a starting point, outlining his findings and how discriminants have been used as a tool in neologism retrieval in the work of numerous researchers, but also how they have been spared little attention even in this regard. Using the *OED* to collect a total of 4000 words introduced to the English language in eight separate years 50 years apart, I found that there was some fluctuation in the use of discriminants before 1810, before there was a steady increase in the frequency of discriminant usage. I found that lexical discriminants, quotation marks, and italics were used to signal new words. I also found that quotation marks were not always the most common discriminant, which did not support my hypothesis, but that they far outpaced the use of lexical discriminants and italics once they were eventually used from the latter half of the nineteenth century. The most frequent lexical discriminant throughout the study was *called,* which I had predicted. In the discussion, I conclude that more in-depth research is required to be able to be certain about the link between word type and the discriminant used to signal it, and that the topic of discriminants is one that could be further analysed in a vast number of ways, with a suggestion that a focus on the link between text type and discriminant types would be a particularly beneficial research route.

REFERENCES

Abel, A. and Stemle, E.W., 2018. On the Detection of Neologism Candidates as Basis for Language Observation and Lexicographic Endeavours: The STyrLogism Project. In *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts* (pp. 535-544).

Anderson, J. M. and K. Malmkjær . 2006. The Linguistics Encyclopedia. London: Routledge.

Ayto, J. 1996. Lexical innovation: neologisms and dictionaries. In G. Anderman and M. Rodgers, eds. *Words, words, words: The translator and the language learner.* Clevedon: Multilingual Matters. pp. 63-68.

Baayen, R. H., Tomaschek, F., Gahl, S., & Ramscar, M. 2017. The Ecclesiastes principle in language change. *The changing English language: Psycholinguistic perspectives*, 21-48.

Breen, J. 2009. Identification of Neologisms in Japanese by Corpus Analysis. In *Proceedings of eLexicography in the 21st century*. Universite´ catholique de Louvin

Chlebda, W. 1991. *Elementy frazematyki. Wprowadzenie do frazeologii nadawcy*. Opole: WSP.

Durkin, P., 2014. *Borrowed words: A history of loanwords in English*. OUP UK.

Harley, H., 2017. *English words: A linguistic introduction*. John Wiley & Sons.

Kerremans, D., Stegmayr, S. and Schmid, H.J., 2012. The NeoCrawler: Identifying and retrieving neologisms from the internet and monitoring ongoing change. *Current methods in historical semantics*, *73*, p.59.

Lieber, R., 2005. English word-formation processes. *Handbook of word-formation*, pp.375-427.

O'Donovan, R. & O'Neill, M. 2008. A Systematic Approach to the Selection of Neologisms for Inclusion in a Large Monolingual Dictionary. In J. D. E. Bernal (ed.) *Proceedings of the*

13th EURALEX International Congress. Barcelona, Spain: Institut Universitari de Linguistica Aplicada, Universitat Pompeu Fabra, pp. 571–579.

Ološtiak, M. and Rešovská, S., 2021. A dictionary of Slovak neologisms (a project description). *International Journal of Lexicography*, *34*(3), pp.302-314.

Oxford English Dictionary. 2024. Oxford: Oxford University Press.

Paryzek, P., 2008. Comparison of selected methods for the retrieval of neologisms. *Investigationes linguisticae*, *16*, pp.163-181.

Plag, I. 2003. Word formation in English. Cambridge: Cambridge University Press.

Smyk-Bhattacharjee, D. 2006. 'Acceptance as an integral factor in the interpretation of novel words'. *SKASE Journal of Theoretical Linguistics* 3(2): 28-36.