UNIVERSITY OF HELSINKI

Breaking the Mold

A corpus study of numeral+noun phrases in Scottish Gaelic

Emma McKenzie
Master's Thesis
Linguistic Diversity in the Digital Age
Faculty of Arts
University of Helsinki
May 2020

Tiedekunta – Fakultet – Faculty	Koulutusohjelma – Utbildningsprogram – Degree Programme
Faculty of Arts	Linguistic Diversity in the Digital Age

Opintosuunta – Studieinriktning – Study Track
Diversity Linguistics

Tekijä – Författare – Author
Emma McKenzie

Työn nimi – Arbetets titel – Title
Breaking the Mold: A corpus study of numeral+noun phrases in Scottish Gaelic

Työn laji – Arbetets art – Level	Aika – Datum – Month and year	Sivumäärä– Sidoantal – Number of pages
Master's Thesis	May 2020	56 + 7

Tiivistelmä – Referat – Abstract

This project is a corpus-based study on numeral + noun phrases in Scottish Gaelic. The typical pattern in Scottish Gaelic is to use a singular noun after numerals one and two and a plural noun after numerals three through ten. However, there are some nouns that do not follow this expected pattern. These exceptions are called numeratives and there are three different categories of numeratives in Scottish Gaelic: duals, numeratives identical in form to a singular, and numeratives with a form that differs from singular and plural and only used with numerals. This study aims to find which nouns have numerative forms and how their use varies diachronically and between dialects.

While numeratives have been more researched in Welsh and Irish, there is not much research on numeratives in Scottish Gaelic. Ò Maolalaigh (2013) did a more restricted corpus study to find what nouns use singular after numerals three through ten. The past research provides a good comparison for my results and gives me a good foundation to expand on. From the past research, there seems to be a semantic relationship between the kinds of nouns that have numerative forms, so I sort my results into semantic categories as well. I also look at numeratives from the perspective of linguistic complexity since Scottish Gaelic is a minority language with a large proportion of L2 speakers.

This project uses *Corpas na Gàidhlig* (the Corpus of Scottish Gaelic), which is part of the University of Glasgow's Digital Archive of Scottish Gaelic. I search the corpus for numerals two through four to see which nouns use numeratives and how consistently they use them. I also look at how frequently numeratives are used diachronically and how usage varies across dialects. I focus especially on nouns that have a high number of numerative tokens to see if there is a pattern in their usage.

In my results, I found 47 nouns that use a dual form and 105 nouns that use a numerative identical in form to a singular. The overall findings for numerative use are that dual use is decreasing, while use of numeratives identical in form to singular has been increasing since 1900-1949. The semantic category with the most dual tokens is natural pairs. The nouns with numeratives identical in form to singular tend to be nouns frequently used with numerals, such as measurement words.

Avainsanat – Nyckelord – Keywords
Scottish Gaelic, corpus linguistics, numeral phrases, semantics, morphology

Säilytyspaikka – Förvaringställe – Where deposited
Helsinki University Main Library

Muita tietoja – Övriga uppgifter – Additional information

Table of Contents

List of Tables

List of Abbreviations

*	Ungrammatical
DAT	Dative
DEF	Definite
FUT	Future
GEN	Genitive
MASC	Masculine
NMRT	Numerative
PAST	Past Tense
PL	Plural
PRS	Present Tense
SG	Singular
3	Third Person
CnG: #	*Corpas na Gàidhlig:* Text Number

1   Introduction

1.1   Numbers, Nouns, and Numeratives

Most people do not think much about the morphology of a noun after a numeral when talking about things like the number of objects. This is because most nouns follow a specific pattern and we have learned to follow that pattern. In English: *one apple, three dogs, eleven inches*. We use a singular noun for when there is one object and a plural noun when there is more than one object. There is, of course, variation in how the world's languages construct a phrase about the number of objects. In Scottish Gaelic, it is typical to use a singular noun after numbers one and two and a plural noun after numbers three through ten. But this is not the end of the story or this research project would not exist.

In Scottish Gaelic there are exceptions to the pattern in noun form after numerals. These exceptions are called numeratives. This project will focus on two different types of numeratives in Scottish Gaelic. The first type of numerative is nouns that use a dual form, which is a form that must be preceded by the numeral two and differs from the singular and plural forms of the noun. The second type of numerative is the nouns that use a numerative identical in form to a singular when following numbers three through ten instead of the plural form. There is a third kind of numerative in Scottish Gaelic, which is a unique noun form that differs from both the singular and the plural and is only used after numerals. To the best of my knowledge, there is only one such noun in Scottish Gaelic, but I will not focus on it in this project.

The motivation for this study is to uncover dialectal and diachronic information about numeratives in Scottish Gaelic, as well as finding more nouns that use a numerative form. To date, there has been no comprehensive study of which nouns in Scottish Gaelic use forms that vary from the typical pattern and why the variation exists. Could there be a semantic relationship between nouns that have numerative forms? This research hopes to find a pattern for which form of a noun is used after numerals and how that use may be impacted by the semantics of the noun itself, the time period, or the dialect being used. Has the use of numeratives changed over time? Does the frequency of numerative use vary between different dialects?

One topic that propels these questions is linguistic complexity and how that complexity can change over time based on who is speaking a language. Generally, languages maintain complexity unless there is a large portion of the population learning the language as a second

language (McWhorter 2008: 168). The composition of speakers of Scottish Gaelic has changed drastically, especially since the 1950s (Lamb 2002: 16), so it will offer a unique diachronic perspective of variation in the use of numeratives. Because numeratives are exceptions, they cause the description of numeral + noun phrases to be longer, which increases complexity (Dahl 2004: 22). Additionally, the dual adds a different form along with singular and plural, increasing the number of parts in the number system, which also increases complexity when it is measured as the "number of parts in a system" (Miestamo 2008: 24). This research project will look at whether the complexity of ae endangered language changes or stays the same over time in one part of its grammar.

There has not been much research on numeratives in Scottish Gaelic; however, there is one study that is very important for my research. Ó Maolalaigh (2013) looks at what nouns in Scottish Gaelic use a numerative identical in form to a singular with numbers three through ten. Looking through texts from the 20th century from the same corpus I use in this project, Ó Maolalaigh (2013:132) was able to find 47 nouns that did not follow the typical pattern in Scottish Gaelic. In my project, I will expand the search for numeratives to include the 19th and 21st centuries and compare dialects to build off of the numeratives that Ó Maolalaigh (2013) found and compare my results to his.

While there has been little research on numeratives in Scottish Gaelic, there is more existing work on numeratives in Welsh and Irish, which is where the term "numerative" comes from. While Welsh and Irish do have grammatical differences from Scottish Gaelic, there are semantic and categorical similarities between their numeratives and those of Scottish Gaelic. Because of this, it is relevant to compare the research on Welsh and Irish numeratives to my own research on Scottish Gaelic numeratives, especially the Irish numeratives because of how closely related the languages are (see §2.1 for discussion on the relationship between Irish and Scottish Gaelic). Acquaviva (2006) analyzes the known numeratives in Irish and Scottish Gaelic, which he terms as "special plurals" or "irregular plurals" in Irish and "inherent plurals" in Scottish Gaelic. Acquaviva (2006: 9) notes that many of these nouns are related to counting or measurements/measuring, but "a precise semantic characterization remains elusive". Numeratives in Middle and Modern Welsh have also been studied by Nurmio & Willis (2016) and Nurmio (2019) and they found that most numeratives were nouns for humans, farm animals, and natural pairs (duals). The numeratives found in Irish and Welsh will make an interesting comparison to the numeratives I find in my study to see if there is a semantic pattern across these languages.

Across different projects, what I refer to as a numerative, has been called many other terms. This can be confusing to the reader, so I will make sure to define all of the terms I am using and how they may differ from what other authors have used. Nurmio & Willis (2016: 297) define numeratives as special forms of nouns that are used after numerals, such as *tri broder* 'three brothers' (singular *brawd*, plural *brodyr*). This definition of a numerative is particularly fitting for Middle Welsh, which has a small group of nouns with a unique numerative form which differs from both the singular and the plural (see §2.5), as well as having numeratives that are identical in form to the plural and duals (Nurmio & Willis: 2016: 299). Nurmio (2019: 233-234) also uses the term numerative to describe nouns in Irish that use a plural after numerals where a singular is expected, as well as the unique numerative forms. While Scottish Gaelic has fewer unique numeratives than Irish, it does have dual forms and nouns that do not follow the expected pattern after numerals. In this paper, I will call any noun forms that vary from the typical pattern when following numerals numeratives. While I could use Acquaviva's (2006) term of "inherent plurals", I want to emphasize the relationship between the use of these forms and numerals since that is when these forms are used. Also, while many of the numeratives denote measurements or other frequently counted entities such as money, not all of them do, and so I would argue that not all numeratives are inherently plural, as Acquaviva (2006) would say. Because of the similarities in how numeratives are used between Welsh, Irish, and Scottish Gaelic, adopting the terminology "numerative" from Nurmio & Willis (2016) for my project makes results comparable, and using the same terminology will draw attention to the similarities between the languages. I will also modify the term "dual" for this project. Dual is a number category that is not uncommon in the world's languages. It is generally a number category referring to two entities, but in Scottish Gaelic it always requires the numeral two to precede it. In this paper, a dual denotes a numerative used with the numeral two, unless otherwise stated.

This previous work on other Celtic languages forms the background for my research looking for numeratives in Scottish Gaelic and seeing what the semantic patterns are between them, and whether their use varies diachronically and across dialects.

1.2   Research Questions & Overview

This project is a corpus-based study looking at the morphology and semantics of nouns in numeral + noun phrases in Scottish Gaelic. This research will explore the following questions:

(1) What nouns use a numerative identical in form to singular with numerals three and four?
(2) Which nouns have dual forms?
(3) How consistently do the nouns that have a numerative form use the numerative form?
(4) Does the use of numeratives vary between different dialects?
(5) How does noun form use vary diachronically? Is the use of numeratives changing over time?

For this project, I use *Corpas na Gàidhlig* (the Corpus of Scottish Gaelic), which is part of the University of Glasgow's Digital Archive of Scottish Gaelic (DASG). The corpus is discussed in more detail in §3.2. I use the corpus to find numeratives, specifically nouns that use a numerative identical in form to a singular after numerals three and four and nouns that use dual after two. Then I look at how the use of these non-typical forms is different in frequency in different dialects and time periods, focusing especially on nouns with a high number of tokens to try to more clearly see a pattern. I also group all numeratives into semantic categories to see if there is a pattern in what kinds of nouns tend to use non-typical forms.

Following this Introduction chapter, I have a chapter on background information. Chapter 2 introduces Scottish Gaelic, number in language, numeratives, and linguistic complexity. In Chapter 3, I give information on corpus linguistics, the corpus I use, and my methodology for this thesis. Chapter 4 contains my results, presented in tables, with a focus on frequent numeratives. Chapter 5 discusses the results, what the quantitative results mean, and analyzes the patterns that were found. Finally, Chapter 6 draws conclusions from the results and discussion. Chapter 6 also revisits the research questions posed in this Introduction, summarizes the main findings, and proposes some future research possibilities that build off of this project.

2   Background

2.1   Scottish Gaelic

Scottish Gaelic is a Celtic language in the Goidelic branch, which also includes Irish and Manx (Lamb 2002: 6). Around the year 500 AD, settlers from Ireland moved to Scotland and Man, bringing their language with them (Gillies 2009: 230, Broderick 2009: 305). As the contact with Irish in Ireland decreased and the settlers came into contact with other languages, such as Norse and French, Scottish Gaelic and Manx evolved into their own languages (Broderick 2009: 305, MacAulay 1992: 138-139). Scottish Gaelic branched off

from Irish in the Early Modern period, evidenced by a book from the early 1500s which shows a developed dialect of Scottish Gaelic (Gillies 2009: 230). Because of this we know that at least one Scottish Gaelic dialect existed in the early 1500s, even though it was not until after the 1600s when an early form of Modern Scottish Gaelic started being used in literature (Gillies 2009: 230).

Currently, Scottish Gaelic is one of three official languages spoken in Scotland, along with English and Scots. Scottish Gaelic is considered definitely endangered by UNESCO's Atlas of the World's Languages in Danger, which is defined as "children no longer learn the language as mother tongue in the home" (Moseley 2010). Over the last 150 years, the area where Scottish Gaelic is spoken has shrunk back to the north and west coasts of Scotland and the islands (Gillies 2009: 230). There has also been a significant drop in the use of Scottish Gaelic, especially in the second half of the 20th century (Lamb 2002: 16). In the 2011 Scotland Census, 1.7% of the population reported having some ability to speak Gaelic, which is about 87,100 people (National Archive of Scotland 2020). Revitalization work has been making an effort to bring Scottish Gaelic back into registers where it can be used in everyday life and between the 2001 and 2011 census there was a smaller drop in Gaelic speakers than between the 1991 and 2001 census (BBC News 2013). While the future of Gaelic remains to be seen, there is hope that the number of speakers will stop declining and perhaps begin to increase in the future.

The difference between a language and a dialect is often difficult to define, especially when two languages come from a common ancestor. Sometimes classifying languages or dialects as one or the other is based on mutual intelligibility, but it is more often based on political boundaries instead of differences in the dialects/languages. This is the case with Irish and Scottish Gaelic. Many people think of Irish and Scottish Gaelic as separate languages, but because of their similarities Irish and Scottish Gaelic are often compared in research as dialects. For instance, the *Linguistic Atlas and Survey of Irish Dialects* (Wagner & Ó Baoill 1969) includes Scottish Gaelic in its volumes. According to Lamb (2002: 6), Irish and Scottish Gaelic are only spoken of as distinct languages now because the dialects that once bridged them on the dialect continuum have been lost. I think that Scottish Gaelic and Irish are considered separate languages, not because of a lack of a complete dialect continuum, but because Scotland and Ireland are separate countries. Perhaps if Ireland and Scotland were not separate countries, then Irish and Scottish Gaelic would be considered dialects of each other. Because of this, I think it is relevant to this project to compare Scottish Gaelic numeratives to their Irish counterparts.

2.2   Scottish Gaelic Orthography

This project will be dealing with written Scottish Gaelic from the last two centuries, so it is important to note that over this time Gaelic orthography has changed. While there were dictionaries and grammar guides published as early as 1801 (Ross 2016: 54), the writing system was far from consistent. The first Gaelic Orthographic Conventions report was published in 1981, which was designed to set up an orthography that could be used for taking and grading exams for the Scottish Qualifications Authority (Ross 2016: 62). A second edition of the Gaelic Orthographic Conventions report was published in 2005, with a third edition following in 2009 (Ross 2016: 63). This means that there may be some misinterpretations with older texts because they do not have the same writing conventions as the Scottish Gaelic that is written today. Ross (2016:107) notes some common spelling differences, one of which is <eu> and <ia>, which is based on different dialectal pronunciation. With some words, e.g. *deug* or *diag* 'teen (as in thirteen)' (Ross 2016: 108), there is no confusion on what the words mean, but with *ceud* 'hundred' and *ciad* 'first' (Ross 2016: 114), there does start to be a possible mix up. The current standard is to use <eu> in words such as *ceud* 'hundred' (Ross 2016:138). Another spelling difference Ross (2016: 140) discusses is <s> + stop consonants, which are <sb>, <sp>, <sc>, <sg>, <sd>, and <st>. In Scottish Gaelic <sb> and <sp> are pronounced as /sp/; <sg> and <sc> are pronounced as /sk/; and <sd> and <st> are pronounced as /st/ (Ross 2016:140). Because the pronunciation is the same, these differences in spelling are easy to understand. For example in *scillinn*/*sgillinn* 'pence, penny' (Ross 2016:146) both spellings can be found in texts. This is particularly relevant for my project because *scillinn/sgillinn* 'pence, penny' is a numerative in Scottish Gaelic. According to Ross (2016: 169), these differences in spelling are just orthographic differences in how people wrote the words and do not reflect a different pronunciation. In current orthography, <sp>, <sg>, and <st> are most often used (Ross 2016: 169).

The second orthographic convention I will be discussing, and the one that causes the most difficulties, is the use of accents over vowels. In current Scottish Gaelic orthography, a grave accent is used over vowels to indicate a long vowel, but in the past acute accents were also used, as well as accents not being used at all (Ross 2016: 178-179). The indication of vowel length in Scottish Gaelic is important because it can change the word, for instance with *bàta* 'boat' and *bata* 'staff, stick' (Bauer & Robertson 2016). For this project, this is most relevant with *dà* 'two' and *da* 'for, to' (Bauer & Robertson 2016). In most situations, the context of the sentence can clarify the meaning of the word in question. In situations where

the sentence can be interpreted either way, other sentences from the text can be used to look at if an accent is used and then a best possible guess can be made.

One final note about orthography and morphosyntax, which is essential for understanding the Celtic language examples in this project, is knowing about consonant mutations. In certain environments and under certain conditions, the initial consonant in Scottish Gaelic will change, which is called lenition (Lamb 2002: 18). This mutation can be caused by the noun being preceded by a definite article, a preposition, a numeral, among other possibilities (Gillies 2009: 253). An example of this is *craobh* /kruːv/ 'a tree' compared to *a' chraobh* /xruːv/ 'the tree', where the definite article causes lenition of the initial consonant (Lamb 2002: 19). This can be confusing since in orthography, the consonant does not change; instead lenition is shown by adding an *h* as shown in Table 1 (Lamb 2002: 19). For this project, it is most important to know that *aon* 'one' and *dà* 'two' cause lenition in the following noun in Scottish Gaelic (Lamb 2002: 38).

Table 1: Orthography and Pronunciation of Lenited and Un-Lenited Consonants (Table modified from Gillies 2009: 251)

Not-lenited		Lenited	
Orthography	Pronunciation	Orthography	Pronunciation
b	b	bh	v
c	k	ch	x
d	d	dh	ɣ
f	f	fh	Ø
g	g	gh	ɣ
m	m	mh	ṽ
p	p	ph	f
s	s	sh	h
t	t	th	h

As can be seen from Table 1, while the orthography makes it seem like the consonant does not change, the pronunciation does change when lenited. The consonants l, n, and r do not lenite, as well as /sk/, /sm/, /sp/, and /st/ (Gillies 2009: 251). When looking at examples in this study, there may be an *h* in an example that is missing in another one, which is because of the lenition.

2.3   Dialect Divisions

In this study I have followed the dialect division described by Jackson (1968: 67-68), based on the breaking of long /e:/, which divides the dialects into a central group and a peripheral group:

[I]t is possible to say in very broad terms [...] [that] the central dialect covers the Hebrides as far south as Mull and sometimes further, Ross exclusive of the north-east corner, Assynt, Inverness-shire, western Perthshire, and mainland Argyll roughly north of Loch Awe; while the peripheral dialects comprise Caithness and Sutherland exclusive of Assynt, the north-east corner of Ross, Braemar, eastern Perthshire, the rest of mainland Argyll with Kintyre, and Arran. Moray and the adjacent lower region of the Spey, the wide valley of Strathspey from Rothiemurchus to the Moray border, may go with the peripheral dialects, linking up with Braemar and east Perth (Jackson 1968: 67–8).
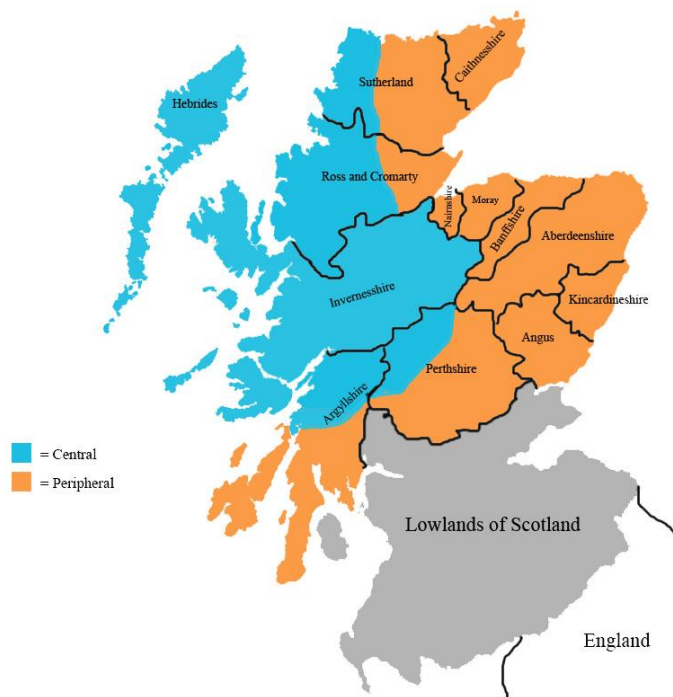
Figure 1: Map of Dialect Groups described by Jackson (1968:67-68)

Jackson's (1968: 67-68) dialect division lines up with the *Gàidhealtachd* ('Gaeldom'), which is the area where Scottish Gaelic is spoken (Gillies 2009: 230). The *Gàidhealtachd* is the Northwestern coasts and islands of Scotland and has been the strongest Gaelic-speaking area for the last 150 years (Gillies 2009: 231). This area correlates with the area of the central dialect (blue), which is described as being "largely innovating and homogenous" by Lamb (2002: 6). Outside of the central dialect is the peripheral dialect (orange). This area is Gaelic-speaking, but does not have as strong of a presence as in the central dialect area. The peripheral dialect is described as "more heterogeneous and fragmented" (Lamb 2002: 6). South of the *Gàidhealtachd* are the Lowlands, where Scots has had a stronger presence than

Gaelic for many hundred years (Gillies 2009: 230). Because of this, the Lowlands are not marked with any Gaelic dialects in Figure 1 and are not divided into counties. In my research, I will also have a third dialect group, which is the Canadian dialect of Scottish Gaelic.

Between 1815 and 1840, 20,000 Scottish Gaelic speakers emigrated from Scotland to Nova Scotia, which was one of the main areas for Scottish immigrants (Nilsen 2010: 94). By 1901, there were an estimated 50,000 Gaelic speakers in Nova Scotia, with many more thousands in other parts of Canada (Nilsen 2010: 101). This was a large enough population that in the late 1800s, Gaelic was the third most spoken language in Canada (Nilsen 2010: 90). Some of the texts that are used in this research are from Canada and due to the distance and probable isolation between Canada and Scotland; the Canadian data will be separated into its own dialect group.

It could be hypothesized that because the Central dialect group is where the level of Gaelic speaking has been the strongest for the longest, that numeratives have stayed consistent in frequency of use. This is looked at in §4.1 and §4.2, specifically in Tables 7 and 12.

2.4   Number in Language

Number is a grammatical category, which marks the number of beings or items involved in an event (Corbett 2000: 2, 4). This distinction in number can be as basic as only having the difference between singular (one entity) and plural (more than one entity), but can also be as complicated as having many more number distinctions, such as dual and trial, which are for two and three beings or items, respectively (Corbett 2000: 1, 4). Corbett (2000: 4) explains:

> […] number, like comparable categories such as gender, case and person, is treated as a 'feature'. This feature is said to have certain 'values' (for number, these include singular and plural, and we have already come across others too). These values of the number feature have meanings and forms associated with them. The main part of the meaning of the singular is that it refers to one real world entity, while the plural refers to more than one distinct real world entity.

The relationship between form and meaning helps us determine the number of the objects in question (Corbett 2000: 4). When comparing languages, the forms may differ, but if the meaning is the same they can easily be compared. A dual means there are two of something and the form varies across languages, but we can still compare the two duals because they

both refer to two entities. A problem arises when the meaning is not exactly the same across languages. If one language has a singular-plural distinction, but another has a singular-dual-plural distinction, both have a plural but the meanings are different (Corbett 2000: 5). In the singular-plural language, the plural is for more than one entity, while for the language with dual, the plural is for more than two entities (Corbett 2000: 5). This requires the linguist to be careful in explaining terms that are used and what their meaning is across languages. For this project, this is less of an issue because the languages I compare to Scottish Gaelic have comparable number systems.

While across languages singular means one and plural means many, there are obviously many different forms that can be used to express these meanings, even within one language. For variation in forms, there are many ways a language can mark number: using numerals, using an article, using an affix on a noun, etc. (Corbett 2000: 4). For example, in English the lack of a marker on a count noun usually indicates that the entity is singular, as in *dog*, while plural has a marker, which is usually *-s*, as is the case in *dogs,* but there are also some irregular markers, as is the case with *woman* vs *women* (Corbett 2000: 4). A word ending in *-s* is not always an indicator for plural either. For nouns like *scissors* and *trousers,* the form appears to be plural, but the meaning is not.

Specifically, this project is looking at numeral + noun phrases. Across languages, there is variation in these phrases. With languages that have a singular-plural opposition, there are three different patterns that emerge with numerals: plural, singular, or both (Corbett 2000: 211). The noun phrase having a plural form is the pattern in English: *three cats*. In Welsh and Irish, the typical pattern is that the noun phrase uses singular (see §2.5). The third pattern, where there is both singular and plural in the noun phrase occurs in Slavonic, based on rules and context, such as case, which can cause the adjective in the noun phrase to be plural while the noun is singular (Corbett 2000: 211). For discussion on number and numeral phrases in Scottish Gaelic, see §2.6.

One concept that Corbett (2000: 95) introduces is a minor number. This is a number category, like singular and plural, but it exists for only a small number of nouns out of all the nouns in a language (Corbett 2000:95). Corbett (2000: 97-99) places four constraints on minor numbers. The first constraint is that the minor number can only exist for a small portion of nouns, out of all the nouns in the language (Corbett 2000: 97). The second constraint is that nouns with minor number forms are a subset of those which have major numbers, which means that every noun with a minor number also has to have all major numbers (Corbett 2000: 99). This means that a noun with a minor dual must have plural and

singular forms as well. The third constraint is that the minor number included in the number system matches "an attested system of number values", which means that there is another language in the world with the same number system (Corbett 2000: 99). For example, having a minor dual is valid because a singular-dual-plural distinction does occur as a major number system, but having a minor trial but no dual, which would make a singular-trial-plural system, would not be valid because no known language has a trial number category without having a dual (Corbett 2000: 39, 99). The final constraint is that if the minor number were removed, the number values would match "an otherwise attested system of number values", which emphasizes that minor numbers are "an 'extra' without which the number system would be normal" (Corbett 2000: 99). An example of a minor number is the dual in Modern Hebrew [Afro-Asiatic]. Unlike in Scottish Gaelic, the dual in Modern Hebrew does not require the numeral two to precede it. In the past, Hebrew had a complete dual system, but it has been lost and now under a dozen nouns still have a dual form (Corbett 2000:95). Because most nouns do not have a dual form in Modern Hebrew, the dual is a minor number in Corbett's (2000: 95) terms.

(1) *ha-yom*   *ʕavar*   *maher*
    DEF-day   pass.PAST.3.SG.MASC   quickly
    'the day passed quickly'   (Corbett 2000:95)

(2) *ha-yom-ayim*   *ʕavru*   *maher*
    DEF-day-DUAL   pass.PAST.3.PL   quickly
    'the two days passed quickly'   (Corbett 2000:95)

(3) *ha-yam-im*   *ʕavru*   *maher*
    DEF-day-PL   pass.PAST.3.PL   quickly
    'the days passed quickly'   (Corbett 2000:95)

Most of the nouns that still have dual forms are measurements for time, like 'day' and 'month', as can be seen in example (2) with 'day' (Corbett 2000: 96). The dual in Modern Hebrew is a minor number since the dual form is only available to a small number of nouns, all of the nouns with a dual also have singular and plural forms, singular-dual-plural is a number system that exists in many languages, and if the dual were removed, Modern Hebrew would have a regular singular-plural number system.

     Maltese [Afro-Asiatic] is another language which Corbett (2000: 96) cites as having a minor number, which also has a dual form that is only retained for a small number of nouns. As with Modern Hebrew, the dual in Maltese does not require a two to precede it and includes only a small number of nouns. There are 32 nouns where a dual form exists, but only eight nouns where use of the dual form is mandatory (Fenech: 1996: 94-95). The 32 nouns

fall into categories such as words for time, weight, measurements, and numerals, and the eight nouns where the dual form is mandatory are all words for time or numerals (Fenech 1996: 94-95). Two examples of nouns with a mandatory dual form are *sena* 'a year', with the dual being *sentejn* and *mija* 'one hundred', with the dual *mitejn* (Fenech 1996: 94). As is the case with Modern Hebrew, with the duals included the number system, Maltese is singular-dual-plural and with the dual removed, the number system is just singular and plural, which are both attested number systems, confirming the dual in Maltese as a minor number (Corbett 2000: 100).

As with Maltese and Hebrew, the Scottish Gaelic dual form is only available for a small group of nouns, as are the numeratives identical in form to a singular. Corbett (2000) does not use the term numeratives, but I believe that numeratives could be a type of minor number since Scottish Gaelic numeratives fit all four of the constraints proposed by Corbett (2000: 97-99). However, Nurmio & Willis (2016: 302) point out that Welsh numeratives (which also applies to Scottish Gaelic numeratives) are not minor numbers because they require a numeral to precede them; they cannot stand alone and represent a numerical value. This is an excellent point and as Nurmio & Willis (2016: 302) decide to use the term "minor category" for numeratives instead of "minor number", I too will classify numeratives as a minor category. This would mean that while numeratives are not exactly what Corbett (2000) would classify as a minor number, they certainly are related.

2.5 Numeratives in Other Languages

The main focus of this project is numeratives, which I define as noun forms after numerals which deviate from the regular agreement pattern of the language. In fact, the term numerative comes from previous work on Welsh by Nurmio & Willis (2016) and was applied to Irish as well. Irish and Welsh are grammatically different from Scottish Gaelic; however there are semantic and categorical similarities across the numeratives of all three languages. The numeratives in Irish are especially valuable to compare to Scottish Gaelic due to the close relation of these two languages, as discussed in §2.1. I will also briefly discuss numeratives in Manx.

While Welsh is not as closely related to Scottish Gaelic as Irish, it still provides valuable background on numeratives. Unlike in Scottish Gaelic, where a plural noun is typical after numbers three through ten, in Welsh it is typical to use a singular noun: *un siop* 'one shop', *tair siop* 'three shops', *pedair siop* 'four shops' (Nurmio & Willis 2016: 298). In

Modern Welsh, the numerative system has mostly been lost, but in Middle Welsh, the words that had numeratives were mainly nouns for humans, for example *brawd* 'brother', plural *brodyr*, and numerative *broder* e.g. *tri broder* 'three brothers' (Nurmio & Willis 2016: 299). Other nouns that have a unique numerative form in Middle Welsh include *blywddyn* 'year', *bu(w)ch* 'cow', *chwaer* 'sister', *dydd* 'day', and *mab* 'boy, son' (Nurmio & Willis 2016: 313). The only word that still has a numerative in Modern Welsh is *blwyddyn* 'year' (Nurmio & Willis 2016: 298). See example (4) below to see how the form for 'year' used after two and three differs from both the singular and plural form of the word. As discussed in §2.2 with Scottish Gaelic, Welsh also has consonant mutations that are triggered by some numerals.

> (4) *Blwyddyn* 'year'
> *Blynyddoedd* 'years'
> *Un flwyddyn* 'one year'
> *Dwy flynedd* 'two years'
> *Tair blynedd* 'three years' (this form applies to 4+ as well)
> (Nurmio &Willis 2016: 298)

Middle Welsh also has numeratives that use a form identical to a plural with numbers three through ten. These nouns include *gŵr* 'man', *gwraig* 'woman', *march* 'horse', *merch* 'girl, daughter', *nant* 'river, stream', and *tŷ* 'house', (Nurmio & Willis 2016: 313). Nurmio & Willis (2016: 316) show that most numeratives in Middle Welsh denote animates and follow the animacy hierarchy. According to the animacy hierarchy, nouns that are higher up in the hierarchy, such as humans and then animals, are more likely have irregular number morphology (Corbett 2009: 6, 55). Minor numbers have been shown to not be conditioned by the animacy hierarchy, while major numbers are, and it may be that minor categories (numeratives) also are not conditioned by the animacy hierarchy (Corbett 2009: 96-97). This makes it interesting to note that the numeratives in Middle Welsh do follow the animacy hierarchy. Nurmio & Willis (2016: 316) also note that while 'year' and 'day' may not be high on the animacy hierarchy, they are nouns with high frequency with numerals specifically, which likely impacts why they have numerative forms.

Out of the Welsh numeratives, it is very interesting that the only measurement-related nouns are 'day' and 'year', which are also nouns with minor number forms in Maltese and Modern Hebrew (see discussion in §2.4). While some of the nouns with numeratives in Welsh are nouns that would often be counted, such as 'cow' and 'man', others do not seem to fit into this semantic category, such as 'river' and 'house'. Welsh adds different semantic categories to compare Scottish Gaelic numeratives to.

Irish is very useful to look at for studying numeratives in Scottish Gaelic. Though there has never been a full corpus search for numeratives in Irish, they have been more researched than in Scottish Gaelic, so not only are numeratives identified, but there is also known dialectal variation (Acquaviva 2006: 1865-1866). In Irish, a singular noun after numbers three through ten is typical. The nouns with numerative forms in Irish tend to be similar to the ones that have been identified in Scottish Gaelic, but they are not exact parallels since there are some grammatical differences between the two languages. The numerative forms in Irish are forms identical to plural after numbers three through ten, unique numerative forms (which Acquaviva (2006) calls "irregular plurals", as mentioned in §1.1), and duals. The following lists of nouns with numeratives may not be exhaustive; they are merely the ones that have been noted in various grammars. A comprehensive corpus study of Irish numeratives is a desideratum.

The largest category of numeratives in Irish are those that use a form identical to plural after numbers three through ten. In all dialects, the numeratives that take a form identical to plural are: *ceann* 'head (as a unit), one', *cloigeann* 'head (counting persons)', *orlach* 'inch', *troigh* 'foot (measure)', *slat* 'rod (measure), yard' (Ó hAnluain 1999: 70). Numeratives in Munster Irish are listed as *nóimint* 'minute', *punt* 'pound', *bád* 'boat(ful)', *turas* 'time, occasion' (Ó Siadhail 1982: 102-104, 1989: 166-168), *bó* 'cow', *cnámh* 'bone', *maidean* 'morning', *méar* 'finger', *naomhóg* 'coracle, small boat', *capall* 'horse', *cos* 'foot', *dosaen* 'dozen', *focal* 'word', *fód* 'strip of land', *piunt* 'pint', *soitheach* 'vessel, jar', *tigh* 'house', *bua* 'life, gift', *suíochán* 'seat', and *gort* 'field' (Ó Sé 2000: 225-228). In the Connacht dialect, the numeratives are listed as *bord* 'edge, side', *ceathrú* 'quarter', *cloch* 'stone (of weight)', *dual* 'strand', *galún* 'gallon', *lámh* 'hand', *mála* 'bag(ful)', *punt* 'pound', *stór* 'storey', *ualach* 'load' (Ó Siadhail 1982: 102-104, 1989: 167-168), *scór* 'score' and *ubh* 'egg' (de Bhaldraithe 1953 [1977]: 7). Finally, in Ulster Irish the numeratives identical to plural are *capall* 'horse', *fód* 'strip of land', *ceann* 'head (as a unit), *cos* 'foot', *doras* 'door', *bascáid* 'basket(ful)', *uan* 'lamb', *méar* 'finger', *seachtain* 'week', *cloch* 'stone (of weight)', *teanga* 'language', *orlach* 'inch', and *slat* 'rod (measure), yard' (Ó Baoill 1999: 104-105). Across the dialects, there is a semantic similarity and overlap with what nouns have numeratives. Most of the numeratives in this category are nouns for measurements or amounts. Munster has the most numeratives with 20 nouns listed here. Ulster has 13 numeratives listed, and Connacht has 12. Munster and Ulster having the most numeratives is not surprising since according to Acquaviva (2006: 1866) in these two dialects, the use of plural after numerals three through ten is more accepted. Munster Irish clearly has more

numeratives in this category, making it the most similar to Scottish Gaelic since the use of plural after number three through ten is more common. However, Connacht does not appear to be too different from Ulster since it only has one less numerative listed here. Of course, these lists are not exhaustive and it is possible that Ulster has more numeratives. I will compare these numeratives to my results and then see if Munster is also the most similar to Scottish Gaelic in what nouns are numeratives.

Irish also has numeratives that have a unique form which differs from the singular and plural forms of the noun. These unique numeratives were actually once the regular plural forms, but probably due to having plural markers that were no longer productive, new plurals were created with the old ones becoming restricted as numeratives (Greene 1992: 527). Additionally, for some nouns there was no difference between plural and singular forms, which caused the pattern of using singular after numbers three through ten to spread to all nouns (Greene 1992: 528). This is why Irish uses a singular noun where Manx and Scottish Gaelic did not apply the pattern to all nouns and still use a plural after numbers three through ten (Green 1992: 527-528). So the unique numerative form is a historic plural, while the currently used plural is a newer form. These words tend to be words dealing with counting units, which seems to be the overall trend with Irish nouns that do not follow the typical pattern after numerals. Though they vary dialectally, the words that have special forms after numbers in all dialects are listed in Table 2 below.

Table 2: Irish Nouns with Numeratives (Table modified from Acquaviva 2006: 1865-1866)

Singular	Plural	Numerative form
bliain 'year'	*blianta*	*bliana*
fiche 'twenty'	*fichidí*	*fichid*
pingin 'penny'	*pinginí*	*pingine*
seachtain 'week'	*seachtainí*	*seachtaine*
scilling 'shilling'	*scillingí*	*scillinge*
uair 'time, occasion'	*uaireanta*	*uaire*

In Table 2, we can see that the form of these nouns after numbers three through ten differs from both the singular and plural forms, making these unique numeratives. Munster Irish also has two nouns listed as having a unique numerative form *feara* 'man' with singular *fear*, plural *fir* and *real(t)a* 'sixpence' with singular *real* and plural *réalacha* (Ó Siadhail 1982: 102, 1989: 166).

Finally, the last numerative category that Irish has is dual. Irish has traces of the old dual system from Old Irish left, especially for natural pairs. These are listed as *cos* 'foot', *bos*

'palm (of the hand)', *bróg* 'shoe', *cluas* 'ear', and *lámh* 'hand' (Christian Brothers 1960: 82). As is the case in Scottish Gaelic, the dual in Irish also requires that the numeral two precede the noun (Greene 1992: 534). The nominative singular form is *lámh* 'hand', while the dual is *dhá láimh* [two hand.DUAL] 'hands (of one person)' and the same pattern exists for other duals (Christian Brothers 1960: 82). In Wagner's (1964, 1966) and Wagner & Ó Baoill's (1969) survey of different Irish dialects, the use of the dual is still widespread, especially in Munster and Connacht dialects. When responding to the translation for 'ears', for example, both the plural and dual form answers were given, but this could be because they are asked to translate from English which has no dual form for 'ears' (Wagner 1964, 1966, Wagner & Ó Baoill 1969). When translating 'ears', most Ulster speakers gave only the plural form (Wagner 1964, 1966, Wagner & Ó Baoill 1969).

Like Scottish Gaelic, Manx also uses a singular noun after numbers one and two and then a plural noun after numbers three through ten (Draskau 2012: 108). Unfortunately, the numeratives in Manx have not been collected. However, Draskau (2012: 109) notes that "Nouns of measure usually remain singular after any numeral". She gives the example *punt* 'pound' and *laa* 'day' (Draskau 2012: 109). Manx seems to have a similar pattern with numeratives as Irish, but this requires further study.

Even with Welsh and Irish typically using singular forms after numbers three through ten, while Scottish Gaelic would typically use a plural form instead, I do think that the numeratives across these languages can be compared. My study will add to the overall picture about what nouns have numerative forms and what semantic categories they belong to. Irish also provides known dialectal variation, so it is possible to compare which dialect has the most overlap with Scottish Gaelic. Of all the numeratives and minor numbers discussed in this paper thus far, 'year' and 'day' seem to be important nouns for these categories. At least one of these nouns has a minor number or numerative in Modern Hebrew, Maltese, Welsh, Irish, and Manx. These nouns could be persisting in minor category use due to their high frequency, especially with numerals. See §2.7 for a further discussion on frequency and how it could impact the retention of the numerative category diachronically.

2.6 Numeratives in Scottish Gaelic

In Scottish Gaelic it is typical to use a singular noun after numbers one and two (examples (5) and (6)) and then a plural noun after numbers three through ten (examples (7)-(9)). It is also typical to use a singular noun with multiples of twenty and one hundred (example (10))

(Lamb 2000: 38). In this project, numeral + noun phrases using numerals 20 and greater are labelled as phrases with complex numbers and separated into their own category, which is discussed further in §3.3.

(5) *aon   leabhar*
    one   book.SG
    'one book'   (CnG: 108)

(6) *dà   leabhar*
    two   book.SG
    'two books'   (CnG: 204)

(7) *tri   leabhraichean*
    three   book.PL
    'three books'   (CnG: 315)

(8) *ceithir leabhraichean*
    four   book.PL
    'four books'   (CnG: 130)

(9) *deich leabhraichean*
    ten   book.PL
    'ten books' (CnG: 400001)

(10) *fichead   leabhar*
     twenty   book.SG
     'twenty books'   (CnG: 238)

Numerals such as *fichead* 'twenty' and *ceud* 'hundred' were historically nouns and were originally followed by a genitive plural (Gillies 2009: 264). However, the genitive plural and nominative singular forms of many common nouns were indistinguishable and the pattern of using a singular noun after these numerals spread to all nouns (Gillies 2009: 264).

In Scottish Gaelic the singular noun is unmarked and plural is marked with a suffix, and/or palatalization. Common suffixes are *-ean*, *-an*, *-achan*, *-ichean*, and *-nnan* (Lamb 2002: 26-27). Some examples include: *compiutair* 'computer' *compiutair-ean* 'computers'; *litir* 'letter' *litr-ichean* 'letter'; and *loch* 'lake' *loch-an* 'lakes' (Lamb 2002: 27). Palatalization is usually marked by an *i* in orthography (Lamb 2002: 20). Examples of this include *fear* 'men', *fir* 'men' and *boireannach* 'woman', *boireannaich* 'women' (Lamb 2002: 26).

As discussed above, the nouns that differ from the typical pattern are numeratives. There are three classes of numeratives in Scottish Gaelic: nouns that use a form identical to singular after numbers three through ten; nouns that have a unique form used with numerals; and nouns that have a dual form used after 'two'. Examples of numeratives are given below.

(11) *trì   bliadhna*
    three   year.NMRT
    'three years'   (CnG: 108)

(12) Example of a dual for *bròg* 'shoe', pl. *brògan*
    a. *an   aon   bhròg   bheag*
      the   one   shoe.SG   little.SG
      'the one little shoe'
    b. *an   dà   bhròig   bhig*
      the   two   shoe.DUAL   little.DUAL
      'the two little shoes'
    c. *na   trì   brògan   beaga*
      the   three   shoe.PL   little.PL
      'the three little shoes'   (Gillies 2009: 255)

(13) *bò* 'cow' *crodh* 'cattle' *còig ba* 'five cows'
    (*crodh* is an uncountable mass)   (Acquaviva 2006: 1864)

Example (11) shows that the word for 'year' uses singular after a numeral instead, making it a numerative. It should be noted that while example (11) uses the numerative form of *bliadhna* 'year', numerative use is not absolute and there are a few occurrences of the plural form (*bliadhnaichean* 'years') after numerals that were found in my results (see Table 10 in §4.2) and Ó Maolalaigh's (2013: 129) results, but these are a small proportion of the total tokens. In (12), we see an example of a noun and an adjective with a dual form. As noted in §1.1, the dual in Scottish Gaelic requires *dà* 'two' to precede it; the dual form of the noun cannot stand alone (Gillies 2009: 264). In example (12) we can see that there is also a distinct dual form of the adjective which differs from the singular and plural forms. In Scottish Gaelic, adjectives agree in case, number, and gender with the noun (Lamb 2002: 39). This project will not be looking at adjectives due to limitations in time and scope, but it would be a very interesting future research project to see what adjectives have dual forms, how frequently they are used, etc. Example (12) also shows a consonant mutation in (12a) and (12b), as discussed in §2.2. Example (13) is the paradigm given by Acquaviva (2006 1864), which to my knowledge is the only unique numerative in Scottish Gaelic. *Crodh* 'cattle' is also used as a translation for the plural 'cows' in Wagner & Ó Baoill's (1969: 212, 218, 249, 264, 273) dialect survey.

    The first type of numerative in Scottish Gaelic is nouns that use a numerative identical in form to singular instead of plural after numbers three through ten. The numeratives in Scottish Gaelic have not been researched in-depth, but Ó Maolalaigh (2013) did a corpus study on 57 texts from *Corpas na Gàidhlig* (the Corpus of Scottish Gaelic) to find what nouns use singular after numbers three through ten and in which semantic categories they

belong in. Before Ó Maolalaigh (2013), Greene (1992: 532) had listed these nouns as: *fichead* 'twenty', *ceud* 'hundred', *mile* 'thousand', *dusan* 'dozen', *duine* 'person', *latha* 'day', *bliadhna* 'year', but he did not state whether he thought this list to be exhaustive. Ó Maolalaigh (2013:132) expanded that knowledge to 47 nouns from his search, which are displayed below in Table 3, organized by the semantic category Ó Maolalaigh (2013:132) gave to them. Some of the nouns fit into more than one category, so they are also listed more than once, such as *sluagh* 'people', which is listed in the People category, as well as in the Collective category (Ó Maolalaigh 2013: 132). For the table of semantic categories for my results, see Table 19 in §5.4.

Table 3: Nouns that use Singular after Numerals Three through Ten (Table modified from Ó Maolalaigh (2013: 132))

Semantic Category	Words
Numeral	*ceàrn* ('quarter'), *cairteal / cairteil* ('quarter'), *ceathramh* ('stanza', lit. 'quarter'), *ceud* ('hundred'), *ceudamh* ('hundredth'), *còigeamh / còigeadh* ('province', lit. 'fifth'), *dusan* ('dozen'), *fichead* ('twenty'), *ficheadamh* ('twentieth'), *mile* ('thousand'), *mìleamh* ('thousandth'), *muillean* ('million'), *naodhnar* ('nine people'), *ochdamh* ('eighth')
Time	*aois* ('age [of animal]'), *bliadhna* ('year'), *cairteal / cairteil* ('quarter'), *cuairt* ('time', lit. 'round'), *latha / là* ('day'), *mionaid* ('minute'), *oidhche* ('night'), *seachdain* ('week'), *tràth* ('time, day'), *uair* ('hour, time')
Length/distance	*aitheamh / faitheamh* ('fathom'), *cairteal / cairteil* ('quarter'), *fad* ('length'), *mile* ('mile'), *slat* ('yard [measurement]'), *troigh* ('foot [of measurement]')
Measure	*bolla* ('boll, bag'), *meud* ('size'), *uiread* (*urrad* 'quantity, amount')
Weight	*dram* ('dram [of weight]'), *tunna / tonna* ('ton'), *ùnnsa* ('ounce')
Currency	*gròt* ('fourpence'), *not*(*a*) ('pound [sterling]'), *peighinn* ('penny'), *punnd* ('pound [currency]'), *sgillinn / sgilling* ('penny'), *tastan / tasdan* ('shilling')
People	*duine* ('man, person'), *pearsa* ('person'), *naodhnar* ('nine people'), *sluagh* ('people')
Collective	*buntàt*(*a*) ('potatoes'), *luingeas* ('ship(s)'), *sluagh* ('people')
Abstract	*beannachd* ('blessing'), *leum* ('a jump'), *mallachd* ('curse'), *seòrsa* ('type')

From Table 3, there seems to be a semantic theme with most of the numeratives being things that would often be counted, such as money, time, and measurements. There is a small group of nouns, such as *beannachd* 'blessing', *mallachd* 'curse', and *leum* 'jump' that do not seem to belong to the category of countable things. It is possible that when there are nouns that do not seem semantically typical for numeratives, that they are part of fixed phrases, so the numerative form is always used with certain numbers. For example, the phrase *guidhe nan*

seachd mallachd 'the wishing of seven curses' (CnG: 35) is mentioned by Ó Maolalaigh (2013: 133). There are also many numeral words that Ó Maolalaigh (2013) included, even though some of them, like *fichead* 'twenty' and *ceud* 'hundred' are regularly expected to be singular in form after numerals three through ten (as seen in example (10)). The larger ordinals, such as *ficheadamh* 'twentieth' and *ceudamh* 'hundredth' are also expected to use a singular form after numerals three through ten because they would be complex numbers (numerals 20 and greater) in these situations as well e.g. *trì ficheadamh* 'sixtieth'. In my results, these words will be separated into a different category.

Ó Maolalaigh (2013: 129) also looks at the frequency of the numerative in comparison to the plural and the frequency of the numerative per numeral for three through ten. As mentioned above, Ó Maolalaigh (2013) counts number words, such as 'twenty' as numeratives, even though they are expected to use a singular form after numerals three through ten. Other than the number terms, he found that *bliadhna* 'year' and *latha/là* 'day' were the two nouns with the highest frequency of numerative use, followed by *duine* 'man' and *sgillinn/sgilling* 'penny' (Ó Maolalaigh 2013: 129). Ó Maolalaigh (2013: 130) also looked at what numerals had the most occurrences of numeratives that were not frequent and found that *trì* 'three' and *seachd* 'seven' were the numerals with the highest, with *ceithir* 'four' and *còig* 'five' as the second highest. I expand on his findings by looking at more texts from a wider time span, as well as seeing how numerative use changes diachronically and across dialects.

The second type of numerative in Scottish Gaelic is the dual. There is some confusion surrounding the dual form in Scottish Gaelic, which stems from the fact that the nominative dual form looks the same as the dative singular form (Lamb 2002: 25).

(14) *aon chas* 'one foot', *dà chois* 'two feet' (Gillies 2009: 264)

(15) *Cha seas   càirdeas air aon   chois*
     Not stand.3.SG.PRS kinship   on   one   foot.DAT.SG
     'Kinship does not stand on one foot'   (CnG: 247)

In example (15), we see that the dative singular has the same form as the dual of 'foot' in example (14). This similarity in form is actually the leftovers of the Old Irish dual form coincidentally looking the same as the dative and not because they are the same thing (Gillies 2009: 264). Specifically, the dual distinction exists only for nouns that in Old Irish were feminine ā-stems, such as *cas* 'foot' (Gillies 2009: 258, 264). The dual forms are also not very common in use and have not been for a long time. Already in the early 1900s, Mackinnon (1911:5) commented that the use of the dual form had, even in literary language,

dwindled to a small group of nouns, noting that "all careful writers" would always use dual forms for *cas* 'foot', *làmh* 'hand', *cluas* 'ear', and *bròg* 'shoe'. Mackinnon (1911: 5) noted that "old people in some districts" used a dual and dative singular form of 'dog' *choin* (nominative singular *cù* 'dog'), but that it had already otherwise disappeared from both everyday speech and literary language. Another number distinction that Mackinnon (1911:5) describes as being lost is *aon chara* 'one friend', *dà charaid* 'two friends', and *tri càirdean* 'three friends', which is now just *aon/dà charaid* 'one/two friend(s)' and *tri càirdean* 'three friends'. In that distinction, it is interesting that the dual form replaced the singular form instead of the dual form being lost completely. Mackinnon (1911: 1, 6) seemed to believe that the nouns that retained their dual form when he was writing would stay stable in their use in spoken Gaelic, even while commenting on the dwindling use amongst the youth. Young speakers in the north were using a singular instead of a dual form more often, and young speakers in the south were using dual more commonly (Mackinnon 1911: 6). This suggests that there may be a dialectal difference in the use of dual, though I separate the dialects in a different way in my study (see §2.3 for a map of how I divide the dialects).

It could be hypothesized that the loss of the dual has been a trait of language death since Scottish Gaelic is an endangered language. However, as seen in §2.4, with Modern Hebrew and Maltese, it is also not uncommon for the dual to be retained by only a small group of nouns in a language that is not endangered. This means that the loss of dual in Scottish Gaelic cannot be pinned completely on the beginnings of language death.

The third and final type of numerative in Scottish Gaelic is nouns with a unique form that is only used with numerals. There is only one example mentioned in the literature, given in example (13) which is for *bò* 'cow'. However, this numerative will not be analyzed in this thesis, due to time and scope. In my corpus searches, I look for unique numerative forms that appear in my searches, in case there are other ones that exist, but I do not focus on trying to find them specifically.

With Modern Hebrew, Maltese, Irish, and some of the Welsh numeratives, as well as the known numeratives in Scottish Gaelic, it seems like the semantic categories that tend to retain minor categories for number are very similar. For all of these languages, there are measurement terms or nouns that are counted or associated with numbers that have retained these forms that differ from the usual pattern of numeral + noun phrases. For Modern Hebrew, the nouns that retained a dual were measurements for time, such as 'day' (Corbett 2000: 95), which is also a numerative that Ó Maolalaigh (2013) found, along with many other nouns for time with numerative forms. In Maltese, the nouns that require a dual form are also

words for time or number such as 'year' and 'one hundred' (Fenech 1996: 94-95), which are also both nouns that Ó Maolalaigh (2013) found to have a numerative, though I disagree with most of the nouns in his Number category since they are complex numbers that normally take the singular form after numerals three through ten. The only numerative that has remained in Modern Welsh is the word for 'year' (Nurmio & Willis 2016: 298), which is also a numerative in Scottish Gaelic (Greene 1992: 532). The numeratives in Middle Welsh are more oriented on humans and farm animals, but Irish numeratives, like duals in Modern Hebrew, tend to be nouns dealing with measurement. For both Irish and Scottish Gaelic the duals are also similar, since they seem to be retained for nouns denoting natural pairs. So while there are some outliers currently with the known Scottish Gaelic numeratives, there does seem to be an overall semantic connection between what kinds of nouns have numeratives. With this project, I proceed to do a more comprehensive corpus search and see if they are also semantically similar to the numeratives and duals in other languages.

2.7   Linguistic Complexity

As Mufwene et al. (2017: 1) express, there has recently been a large amount of interest in linguistic complexity. This is most likely due to the fact that linguistic complexity is a wide-ranging topic and can be applied to many different linguistic features, such as phonemes and cases, as well as different subfields, such as typology and language contact. I decided to look at complexity for this project because I believe that numeratives in Scottish Gaelic can offer an interesting intersection between complexity, number, and language endangerment, especially since numeratives have been previously largely overlooked in this language.

So how is complexity measured? Dahl (2004: 21) explains the complexity of something as being "measured by the length of the shortest possible specification or description of it". This means that the longer the description of an object, the more complex it is (Dahl 2004: 22). In Scottish Gaelic the regular pattern for noun forms in numeral + noun phrases, which most nouns follow, is that there are singular forms of nouns after one, two, and numbers twenty and above and plural forms of nouns for numbers three through ten. This overall is not a very complex system, but the numeratives increase the complexity because they are exceptions. They increase the length of the description of numeral + noun phrases in Scottish Gaelic, if one were to explain them. Miestamo (2008: 24) breaks complexity further down into two separate types of complexity: absolute and relative. Absolute complexity is theory-based, objective, and defines complexity as the "number of parts in a system"

(Miestamo 2008: 24). Miestamo (2008: 24) gives the example of the size of a phoneme inventory, which is a more complex system with more phonemes (which are parts of the system) and less complex with fewer phonemes. Absolute complexity is how Dahl (2004) approaches complexity (Miestamo 2008: 24). Relative complexity is more focused on the speaker of the language and defines complexity as the "cost and difficulty to language users, i.e., how difficult a phenomenon is to process (encode/decode) or learn" (Miestamo 2008: 24). One of the problems with relative complexity is that there are many different perspectives of a language based on if the person acquired the language as an L1 or learned it as an L2 and whether they are the speaker or the addressee (Miestamo 2008: 26). For example, with numeratives it is generally easy for the hearer to understand because they do not require the typical number and noun agreement to be able to follow the meaning of the phrase. The second problem with relative complexity that Miestamo (2008: 26) discusses is the ability to actually measure how costly different parts of a language are for users, which requires studying acquisition and learning by different language users. Because of this, Miestamo (2008: 26-27) recommends using absolute complexity when studying complexity in languages. For this project, I will focus on absolute complexity because of the problems that Miestamo (2008) mentioned with relative complexity. I do think it is important to keep relative complexity in mind however because the speaker population for Scottish Gaelic has changed so much. As discussed in §2.1, virtually no children are acquiring Scottish Gaelic as an L1 (Moseley 2010), but since this project is looking at numerative use diachronically, there are time periods included in my study when Scottish Gaelic was still being acquired as an L1.

The language user makeup of Scottish Gaelic is also important for how the complexity has changed over time. The main factor in a language losing complexity is L2 acquisition by a large portion of the language's users (McWhorter 2008: 168, Karlsson et al. 2008: viii). It is often the case with minority languages and languages getting revitalized that, especially at the beginning, the people who are learning and using the language are L2 learners. This means that it is likely that Scottish Gaelic is losing complexity over time.

Another concept related to complexity is levelling. Levelling is a "paradigm-internal analogy" and an analogy is "any change due to the influence of one form on another" (Joseph 1998: 358). Essentially, levelling is an overgeneralization of a form that then becomes more universal. An example of levelling is the superseding of the *-en* plural ending in English (still found with some nouns, such as *ox* and *oxen*) by the spread of the plural ending *-s* (Joseph 1998: 358). Levelling is one way a language can reduce complexity since before levelling

there is more variation in forms. However, some words and forms are more likely to get levelled than others. The model that Bybee (1995: 429) suggests is that high frequency items are learned independently, while lower frequency items are learned via rules. So as long as a word is high frequency, it is less likely to level from an irregular form to a regular one because the irregular form is learned unrelated to other forms. This is the reason why the English word *man* has retained an irregular plural *men*, since it is so frequently used. It is also important to think about the difference between absolute frequency and relative frequency. Absolute frequency refers to the overall frequency of a token in the data (i.e. how often a word occurs); while relative frequency refers to the frequency of different forms relative to one another, or the frequency with which the token occurs in certain environments (i.e. how many plural tokens occur vs how many singular tokens occur of the same noun or how often each form occurs with numerals) (Corbett et al. 2001: 202-203). For numeratives, this is important because some nouns may occur frequently overall in the language, while some nouns occur more frequently with numerals specifically than other nouns do. The high frequency of some nouns specifically with numerals could lead to learners acquiring numeratives as phrases instead of learning a rule about form after numerals and learning the numeratives as exceptions. For Scottish Gaelic, I would hypothesize that over time, the use of numeratives may decrease because of levelling, especially if the noun with the numerative form is not very commonly used with numerals. On the other hand, I would also hypothesize that for nouns that are frequently used with numerals, the use of numeratives may remain consistent since the form would occur frequently enough to be acquired by learners.

There are many reasons why I hypothesize that the use of numeratives may have decreased in use over time. The first reason is that numeratives add complexity: describing numeral + noun phrases becomes longer when adding in dual, which nouns have a distinct dual form, and which nouns use a numerative identical in form to a singular. The second reason is that there has been an increase in the number of L2 learners of Scottish Gaelic and a decrease in L1 speakers over time. Languages tend to lose complexity over time, but this is counteracted by L1 speakers creating new complexity (McWhorter 2008:168). Since there has been an increase in L2 speakers of Scottish Gaelic, there is less new complexity being added and so I would hypothesize that numeratives are being used less. Finally, the third reason that I think numerative use has decreased is that I believe non-frequent numeratives may have been levelled out in favor of the regular system. However, I do hypothesize that some numeratives will persist over time, as we saw in §2.5 with many numeratives being nouns for measurement, time, and money. These nouns occur frequently enough with

numerals that L2 learners could learn them as part of noun phrases with numerals instead of learning them as exceptions to a rule. So while I hypothesize that overall numerative use is decreasing over time, I believe that frequently used nouns will retain their numerative use.

2.8 Summary

After reviewing the existing literature, there is room for more research to be done on numeratives in Scottish Gaelic. Are there nouns that are not natural pairs that still have a dual form in use, and how frequently are they used? Are there more nouns that use a numerative identical to singular in form after numbers three through ten other than the ones found by Ó Maolalaigh (2013)? Does use of these non-typical forms vary based on dialect? Has the use of numeratives changed in use over time? The numeratives in Scottish Gaelic have potential to be studied much more and could reveal interesting patterns about numeral +noun phrases and linguistic complexity, while also adding comparable results for the other Celtic languages.

3 Data & Methods

3.1 Corpus Linguistics

Simply put, corpus linguistics is using a large collection of texts to study language. More specifically, McEnery & Hardie (2011: 1) define modern corpus linguistics, "as dealing with some set of machine-readable texts which is deemed an appropriate basis on which to study a specific set of research questions". It is important that a computer is able to read the texts because such a large amount of texts is used that searching for the topic of research would take far too long to do manually (McEnery & Hardie 2011: 2). A corpus may be assembled from any form of language, such as written language from newspaper articles or spoken language which has been transcribed from interviews (McEnery & Hardie 2011: 3-4). McEnery & Hardie (2011: 2) also emphasize the importance of using a corpus that contains many texts and many types of text to more accurately make observations about language. Some corpora are annotated, for example marking parts of speech so that one could look up i.e. second singular pronouns or numerals and not have to sort through words that have similar or the same orthography but a different meaning (McEnery & Hardie 2011: 13). For this project, if a corpus were annotated for numerals, then the search results would not be a mix of *dà* 'two' spelled without an accent marker and *da* 'for, to' and would just contain

sentences using *dà* 'two'. Corpora may be used in a variety of ways, but the most relevant approach for this study is using corpora to look at variation diachronically. A corpus may be used to compare texts from different time periods to track gradual changes in the language (McEnery & Hardie 2011: 94). Using a corpus with texts from the past to modern times, a researcher can track how one word may rise in use to gradually replace another or how one form of a word may gradually disappear from use (McEnery & Hardie 2011: 95).

The downside to corpora is that they are most often based on written texts, since historically we are limited to what materials we have and that is usually texts. Especially when looking at data from hundreds of years ago, we lack any direct data on the spoken language. This can be an issue since for some languages, a more formal register is usually used for written materials, which can vary greatly from the spoken language that is used at the same point in time. Because of this, it can be hard to know what spoken, more casual language was like in the past and a full picture of language use is not possible. For more information on the types of texts in the corpus used in this study, see the following section: §3.2.

In terms of corpora on languages like Scottish Gaelic, the situation has recently improved. In the introduction of Anderson's (2013: 13) volume on corpora of the languages of Scotland, she comments that the volume "comes at a time when new corpus resources for the languages of Scotland are making it possible to ask new research questions of these languages". I agree wholeheartedly with Anderson, and I would also add that new corpora open the door to new researchers just starting out on their journeys. Anderson (2013: 14) also emphasized that these corpora are widely available, which Ó Maolalaigh (2013: 113) also comments on in his chapter. Ó Maolalaigh (2013: 113) observes that while English has many corpora and Irish has a handful, there is a lack of publicly available, online corpora for Scottish Gaelic, but this is changing. One could ask how much research has not been done because of this lack of corpora in Scottish Gaelic. In terms of the subject of this project, numeratives, I do not think that sparse research can be completely blamed on the lack of a large, publicly available corpus, but it certainly makes this kind of research much easier. Many researchers could be discouraged from asking certain research questions if they are required to assemble their own corpus before even starting to try to answer them. Blossoming corpora for the languages of Scotland allows "reliable quantification, easy verification, and the identification of linguistic patterning to which intuition or analysis of isolated texts cannot give access" (Anderson 2013: 17). Because of these new corpora, there is plenty of possibility for new research, such as this project, to be done.

3.2 Corpus Background

For my research, I used *Corpas na Gàidhlig* (the Corpus of Scottish Gaelic), which is part of the University of Glasgow's Digital Archive of Scottish Gaelic (DASG). *Corpas na Gàidhlig* has 549 texts from various genres (religious texts, prose, folklore, songs, etc.) from the mid-1600s to 2016. While the data is mainly written text (there is one transcribed conversation in the corpus), the goal of DASG was to get as many registers and genres as possible, so the texts vary from more formal, religious texts and literature to more everyday language with personal diaries and novels written with dialogue (Ó Maolalaigh 2013: 116). One example of a novel with more everyday speech is *Am Fear Meadhanach* 'The Middle Man' (CnG: 11), which is described as having "colloquial, Lewis (Ness) Gaelic from the mid-to late twentieth century", with "examples of direct speech" (*Corpas na Gàidhlig*). Having a broad range of registers is useful for this project so that a pattern within the language in general can be found, as opposed to having to focus only on language use in religious texts or poetry. That being said, we can never truly know how any language was spoken hundreds of years ago since we only have written text as evidence. *Corpas na Gàidhlig* does aim to include speech in the future (Ó Maolalaigh 2013: 116), which, if a similar study were to be done later, would provide an interesting comparison for this study. Currently, DASG does have an audio archive of 140 hours of recordings that are transcribed. Most of the audio recordings were used for collecting material for the Historical Dictionary of Scottish Gaelic. While this is a wonderful source, the *Corpas na Gàidhlig* has a larger amount of material and has more numerals in use in the materials.

The DASG is fairly new, being founded in 2008 and having texts added through 2016. This work is exciting because quantitative analysis is now possible, but there are some drawbacks, such as this corpus not being annotated. Because of this, the possible methods to do an analysis of this type are limited, which is why the method described in §3.3 was chosen. Additionally, because this research project is focused on analyzing numerative use across dialects and time periods, corpus entries without a geographic origin or a date for the language were not included in analysis.

One useful feature of *Corpas na Gàidhlig* is that it distinguishes between the Date of Edition and the Date of Language. Because written language is more conservative than spoken language, the language used in a text may reflect language that was used earlier than when the text was published. The Date of Language, if given, can be as small as a fifty year window, such as 1900-1949 or it can be a whole century if a fifty-year time period cannot be

distinguished. This can be determined based on features such as common orthography for what is known about the language of that time period, like the use of accent markers. While *Corpas na Gàidhlig* contains texts that are from the 1600s, no texts from before 1800 had the required 50-year block for the Date of Language as well as a Geographic Origin, so they were not used. All results are therefore from 1800-2016.

3.3   Methodology

I started by searching the corpus for nouns in numeral phrases using the written forms of the numerals, *dà* 'two', *trì* 'three', and *ceithir* 'four', as well as numeric forms '2', '3', and '4'. I did not continue with numerals after four because of the time limit and scope of this master's thesis. The results for three and four were combined since the nouns that use numeratives with these numerals belong to the same category, which are identical to singular (as explained in §2.6), while the results for two are analyzed in their own category since I am looking for dual forms. The complete list of search terms is listed below, in Table 4:

Table 4: Search Terms

Search Terms Used
d[h,][à,a,á]
t[h,]r[ì,i,í]
c[h,][e,è,é][i,ì,í]th[i,ì,í]r
2
3
4

For each search, I went through the corpus and collected all noun phrases with numerals into a spreadsheet. I made note of the source they came from, the date of the language, and the geographic origin of the text. All tokens were collected whether they used a numerative or not, so that I could look at the frequency of use of the numerative vs the plural for each word. Some nouns may always use a numerative, while some may only use it e.g. 30% of the time, as discussed in §2.5. For the written out search terms, *dà* 'two', *trì* 'three' and *ceithir* 'four', the searches had lenition allowed (since it impacts orthography, as explained in §2.3) and were accent insensitive, to try to catch as many different spellings as possible. When looking at the tokens for the searches for *dà* 'two' and '2', I needed to double check that the duals I found were truly duals and not dative singulars, as discussed in §2.5. To do this, I collected the whole sentence and analyzed all the potential duals to determine if the

form used was a dual or not. For the searches for three and four, it was more straightforward since I just needed to determine if a numerative or plural was being used.

Many of the tokens in my results are only used once with a numeral and that one time is a numerative form, so not much analysis can be done on those nouns. There are also tokens that show up ten times, but only one token uses a numerative and those nouns are also difficult to analyze in terms of numerative use. If a noun had more than ten tokens with a numerative form and the numerative forms accounted for more than 50% of the total tokens, I looked at the noun more closely to see if there was a change in how it was being used across time periods and dialects. Nouns that used a numerative for all of the tokens I found were not analyzed because a 100% numerative use in my data would indicate that there has been no change in numerative use and I am interested in what has changed in the use of numeratives. The nouns that do not consistently use a numerative or use a numerative all the time will be listed but not analyzed in-depth.

In addition to frequency of numerative use for each noun, diachronic variation was also controlled for. As explained in §3.2, *Corpas na Gàidhlig* uses 50-year blocks as the smallest Date of Language measurement. For the sake of having more precise time periods and to be able to see diachronic patterns more clearly, only texts with a fifty year block were used for analysis. I did not use texts that had no Date of Language because I wanted to focus on how numerative use has changed over time. As mentioned in *§3.2,* while the corpus does contain texts from the 1600s, none of the texts from earlier than 1800 have both a geographic origin and a date of language, so they were not analyzed.

Dialectal variation was also controlled for in this project. As mentioned in §2.7, the dialect groups will be split, according to Jackson's (1968: 67-68) groupings, into Central, Peripheral, and then Canada. This grouping calls for Argyll, Perthshire, and Sutherland to be split, but the corpus does not show where exactly the text is from within the counties, so Argyll was put into the Central group and Perthshire and Sutherland were put into the Peripheral group. These were assigned due to location, since Argyll is on the west coast, Perthshire is more central, and Sutherland only has a small part in the Central dialect group (see map in §2.3). The complete list of the dialect groupings with the areas that were assigned to them are listed below:

Aberdeenshire - Peripheral	Lewis - Central
Argyll - Central	Lochaber - Central
Arran - Peripheral	Mull - Central
Badenoch - Central	Perthshire - Peripheral
Barra - Peripheral	Uist (all locations of Uist) - Central

Caithness - Peripheral	Raasay - Central
Coll - Peripheral	Skye - Central
Easter Ross - Peripheral	Strathspey - Peripheral
Eigg - Central	Sutherland - Peripheral
Gairloch - Central	Taransay - Central
Harris - Central	Tiree - Central
Highlands - Central	Wester Ross - Central
Islay - Peripheral	Western Islands - Central
Jura - Peripheral	

I decided to look at numerative use across dialects as well because the speaker composition for these three dialect areas is very different. As discussed in §2.1, the area where Scottish Gaelic is spoken has stayed the strongest on the north and west coasts of Scotland, which is the Central dialect. How the Central dialect area uses numeratives is different from Scottish Gaelic speakers in Canada, whose ancestors left Scotland hundreds of years ago and the Peripheral dialect area, which is an area where Gaelic is spoken less. There appears to be variation in the use of numeratives in Scottish Gaelic and I wanted to see how the different dialect areas use numeratives. To sum up, frequency of use for nouns that use a numerative was analyzed by word, by dialect, by time period and by consistency of use of numeratives (as opposed to using a plural form most of the time).

Significance values were tested with Fisher's Exact Test because often at least one sample size was fewer than five tokens. When all sample sizes were greater than five, then a chi-square test was done as well. Significance tests were done (Tables 6-9 and 11-17) in order to determine if there is a correlation between numerative use and dialect as well as numerative use and time period.

Finding nouns that have unique numerative forms is more difficult than the other two numerative categories. My hope was that while looking for nouns that use duals and numeratives identical in form to singular, some nouns that have a unique numerative form would be found. Unfortunately, in my searches I did not come across any unique numerative forms.

The search results were also divided into categories: Standard, Complex Number, and Or Phrase. The Standard category is for the phrases that had a numeral followed by a noun. While Or Phrase and Complex number tokens were collected to look for interesting patterns, they will not be analyzed in-depth in this project. The main focus will be on the tokens in the Standard category, which use numeratives in simple numeral + noun phrases, found in Tables 5-17 in the Results chapter.

The Complex Number category contains phrases that have complex numbers that included the searched numeral, like sixty (in Gaelic, 'three twenties') or three hundred. As discussed in §2.6, these numbers have a different form than simple numbers, as seen below in example (16):

(16) *trì*   *fichead*   *bliadhna*
     three   twenty   year.SG
     'sixty years'   (CnG: 4)

Tokens with complex numbers were collected into their own category to see how the second numeral word (such as *fichead* 'twenty' in example (16)) and then the following non-numeral noun (*bliadhna* 'year' in example (16)), are impacted by the numeral. However, these tokens were not analyzed further in my results. The results that do not follow the typical pattern as shown above in example (16) are listed in the Appendix.

The final category is for Or Phrases, which are phrases that used two numerals connected by "or", for example: "There were two or three dogs".

(17) *trì*   *no*   *ceithir* *bràithrean*
     three   or   four   brother.PL
     'three or four brothers'   (CnG: 197)

These were collected into their own category to look at if the noun is impacted by the first or the second number in the phrase. Or Phrases are also not included in my results and analysis, but the tokens that vary from the typical pattern, which is shown in example (17), are in the Appendix.

4   Results & Analysis

This chapter presents the results of my corpus study and uses them to explore the research questions stated in §1.2, namely what nouns have numeratives and whether diachrony and dialect have an impact on the use of numerative forms. First I look at what nouns have duals and their usage by dialect and time period, then I move on to what nouns have numeratives that are identical in form to a singular and their usage by dialect and time period.

As mentioned in §3.2 and §3.3, none of the texts from earlier than 1800 have both a geographic origin and a date of language so tables in this section range only from 1800-2016. All of the time periods are in 50-year groups, as given by the corpus. The 2000-2016 group is not a 50-year group since we are still in that time period, so it is representative of a smaller amount of time than the other time periods.

4.1 Dual

This section presents the results of my corpus searches for *dà* 'two' and '2'. There is a total of 4,703 tokens for these searches, split into the three different categories: Standard, Or Phrase, or Complex Number. For a breakdown of the number of tokens per category, see Table 21 in the Appendix. The following table contains the nouns that are found using a dual form, how many tokens use a dual form, how many total tokens there are, and what percentage of the total tokens use a dual form.

Table 5: Nouns that use Dual for d[h,][à,a,á] and '2' Sorted by the Number of Dual Tokens from Highest to Lowest. (All translations are from Bauer & Robertson (2016))

Noun	# of Dual Tokens	# of Total Tokens	% of Dual Tokens
làmh 'hand'	58	76	76.32%
cluas 'ear'	22	22	100%
cas 'foot (body part), leg'	16	26	61.54%
bas 'palm (of a hand)'	8	8	100%
buidheann 'band, faction, group'	7	26	26.92%
clach 'rock, stone'	7	10	70%
slinnean 'shoulder blade'	5	12	41.67%
glùn 'knee'	4	20	20%
long 'ship'	4	17	23.53%
sluagh 'folk, people'	4	10	40%
bròg 'shoe'	3	7	42.86%
cànan 'language'	3	24	12.50%
craiceann 'skin, parchment'	3	4	75%
searrag 'flask, bottle'	3	3	100%
ceatharna 'troop'	2	2	100%
earrann '(part of an) area, piece'	2	6	33.33%
eathar 'small boat/vessel'	2	11	18.18%
fiacal 'tooth'	2	5	40%
òirleach 'inch (measure)'	2	9	22.22%
òrdag 'thumb'	2	3	66.67%
ràmh 'oar, paddle'	2	14	14.29%
tòn 'backside, bum'	2	2	100%
adhrac 'horn'	1	1	100%
aodann 'face'	1	2	50%
basgaid 'basket'	1	2	50%

bean 'woman, wife'[1]	1	2	50%
boiceann 'goatskin'	1	1	100%
caileag 'girl'	1	1	100%
cearc 'hen, chicken'	1	4	25%
cliathach 'side (of a being)'	1	1	100%
coilleag 'shoot, sprout'	1	1	100%
corrag 'finger, toe'	1	1	100%
craobh 'tree'[2]	1	5	20%
cromag 'hook'	1	1	100%
dìnnear 'dinner'	1	1	100%
eilean 'isle, island'	1	9	11.11%
faobhar 'edge (of a blade)'	1	3	33.33%
focal 'word'	1	4	25%
glac 'glade'	1	2	50%
lurgann 'shaft, shin, leg'	1	1	100%
oifigear 'officer'	1	1	100%
piuthar 'sister'	1	13	7.69%
rathad 'road'	1	10	10%
spàg 'paw, limb (of an animal)'	1	1	100%
teaghlach 'family'	1	9	11.11%
teud 'cord, string'	1	1	100%
tulach 'hillock'	1	2	50%

As shown above in Table 5, in the data there are 47 nouns that are found using a dual form after the numeral two. Of those 47 nouns, 25 only use the dual form once. The four nouns with the most dual tokens (*làmh* 'hand', *cluas* 'ear', *cas* 'foot (body part), leg', and *bas* 'palm (of the hand)'), are all parts of the body that come in pairs. Looking through Table 5, there are quite a few other nouns for parts of the body, both human and animal, such as *glùn* 'knee', *òrdag* 'thumb', and *adhrac* 'horn'. Table 5 also gives a good display over the frequency of use of a dual form, with some nouns using a dual form 100% of the time, such as *cluas* 'ear' using a dual form for all 22 tokens found in the corpus, while others, such as *glùn* 'knee', use a dual for only four of the total 20 tokens, which is 20% of the time. This could indicate that the noun determines the frequency of use of dual, but I also want to see if the date of language or dialect has an effect, which can be seen below.

---

[1] I am counting *bean-uasal* 'gentlewoman, lady' as a token for *bean* 'woman, wife'
[2] I am counting *craobh-ùbhlan* 'apple tree' as a token for *craobh* 'tree'
These compound nouns have the same noun as the base as the non-compound version, so the tokens were combined.

Table 6: Dual across Date of Language

Date of Language	# of Dual Tokens	# of Total Tokens	% of Dual Tokens
1800-1849	32	691	4.63%
1850-1899	14	336	4.17%
1900-1949	75	924	8.12%
1950-1999	62	1460	4.25%
2000-2016	4	227	1.76%

For the date of language, 1900-1949 had the highest percentage of dual tokens used at 8.12%. The 2000-2016 group had the lowest percentage of dual tokens at 1.76%, but it was also the least represented in the data with only 227 tokens total. The difference between 1900-1949 and 2000-2016 is statistically significant (Fisher's $p = 0.0002$), even though there is a large difference in total number of tokens. The date of language time period that had the highest number of total tokens was 1950-1999 with 1,460 total tokens, but the percent of dual tokens was 4.25%, which was similar to 1800-1849 and 1850-1899 which had 4.63% and 4.17%, respectively. The difference between 1800-1849 and 1850-1899 is not statistically significant (Chi-squared $\chi2=0.1139$, df = 1, $p = 0.735735$ and Fisher's $p = 0.8725$), while the difference between 1850-1899 and 1900-1949 is statistically significant, which means that the increase in use of dual is significant (Chi-squared $\chi2=5.857$, df=1, $p = 0.015515$ and Fisher's $p = 0.0175$). The difference between 1900-1949 and 1950-1999 is also significant (Chi-squared $\chi2=15.6494$, df = 1, $p = 0.000076$ and Fisher's $p = 0.0001$), but the difference between 1950-1999 and 2000-2016 is not statistically significant (Fisher's $p = 0.0948$). Because many of the differences were significant, this indicates that the date of language is a variable that impacts the percentage of dual tokens used. The increase in the number of dual tokens in 1900-1949 from 1850-1899 was significant and the decrease to 2000-2016 from 1900-1949 was also significant. These results are discussed further in §5.1.

Table 7: Duals across Dialects

Dialect Group	# of Dual tokens	# of Total Tokens	% of Dual Tokens
Canada	0	7	0%
Central	177	3333	5.31%
Peripheral	11	298	3.69%

Looking at dual use across dialects, Canada was underrepresented and only had a total of 7 tokens, none of which used a dual form. While it is possible that Canadian Gaelic has lost dual forms, due to the scarcity of Canadian data, nothing definitive can be said about the use of dual in the Canadian dialect. The Peripheral dialect has 298 tokens, 11 (3.69%) using a dual form. The Central dialect had over ten times the amount of total tokens than the

Peripheral dialect, with 3,333 total tokens. Out of those, 177 (5.31%) used dual. The difference between the Central and Peripheral dialects is not statistically significant (Chi-square $\chi2$=1.4609, df = 1, p = 0.22679 and Fisher's exact test p = 0.2746). The data does indicate that the Central dialect uses dual forms more often, but overall my data does not indicate a statistically significant dialectal difference. However, it would be useful to look at a larger set of data to see whether this finding holds.

Table 8: Tokens of Dual for *làmh* 'hand' over Total Tokens by Dialect and Date of Language

Dialect/Date of Language	1800-1849	1850-1899	1900-1949	1950-1999	2000-2016
Canada	0	0	0	0	0
Central	14/16 [87.5%]	0/3	21/26 [80.77%]	20/27 [74.07%]	0/1
Peripheral	1/1 [100%]	0	0	2/2 [100%]	0

The Canadian and Peripheral dialects have too few tokens to analyze in depth, so I will focus on the Central dialect. As seen in Table 8, for the Central dialect *làmh* 'hand' has 14 dual tokens out of 16 in 1800-1849, which is 87.5%; which decreases to zero dual tokens in 1850-1899. The difference between these two time periods is statistically significant (Fisher's p = 0.0103). In 1900-1949 there are 21 dual tokens out of 26 total *làmh* 'hand' tokens, which is 80.77%, which is also a statistically significant difference from 1850-1899 (Fisher's p = 0.0153). The difference between 1800-1849 and 1900-1949 however is not statistically significant (Fisher's p = 0.6897). In 1950-1999 the percentage of dual tokens decreases to 74.07% and the difference between 1900-1949 and 1950-1999 is not statistically significant (Chi-squared $\chi2$=0.339, df=1, p = 0.560421 and Fisher's p = 0.7445). Finally, in 2000-2016, there are 0 dual token and only 1 use of *làmh* 'hand' total and this is not statistically significant vs 1950-1999 (Fisher's p = 0.2857). Based on this data, the date of language has had an impact on the decrease in use of the dual form of *làmh* 'hand' in the 19th century, as well as having an impact on the increase from 1850-1899 to 1900-1949. While the percentage of use is decreasing overall from 1900-1949 to 2000-2016, these differences are not statistically significant, so based on my data there does not seem to be a statistically significant relation between the dual use of *làmh* 'hand' and the time period in the 20th century.

Table 9: Tokens of Dual for *cas* 'foot (body part), leg' by Dialect and Date of Language

Dialect/Date of Language	1800-1849	1850-1899	1900-1949	1950-1999	2000-2016
Canada	0	0	0	0	0
Central	2/3 [66.67%]	0	5/5 [100%]	7/14 [50%]	1/2 [50%]
Peripheral	0	0	1/1 [100%]	1/1 [100%]	0

Once again, the Canadian and Peripheral dialects have too few tokens so they will not be analyzed in depth. In 1800-1849 for the Central dialect, there are two dual tokens out of three total *cas* 'foot, leg' tokens, which is 66.67%; in 1850-1899 there are no tokens at all, and the difference between 1800-1849 and 1850-1899 is not statistically significant (Fisher's p = 1). In 1900-1949 all five tokens use dual and the difference between 1850-1899 and 1900-1949 is not statically significant (Fisher's p = 1). The decrease between 1900-1949 and 1950-1999 is also not statistically significant (Fisher's p = 0.106). In 2000-2016, there were only two tokens total and the difference between 1950-1999 vs 2000-2016 is not statistically significant (Fisher's p = 0.2857).

The tokens for *cas* 'foot, leg', do appear to show a change over time overall from 66.6% to 0%, then shows an increase to 100% , and finally a decrease to 50% but none of the changes between the time periods are statistically significant. Due to these limitations of my data, no conclusion can be made about whether the dual use of *cas* 'foot' has increased or decreased significantly with time, though the data appears to show an overall decrease in use.

4.2 Numeratives Identical in Form to Singular

This section presents the results of my corpus searches for *trì* 'three', *ceithir* 'four', '3', and '4'. There is a total of 3,088 tokens for these searches, split into the three different categories: Standard, Or Phrase, or Complex Number. For a breakdown of the number of tokens per category, see Table 22 in the Appendix. The following table contains the nouns that are found using a numerative form, how many tokens use a numerative, how many total tokens there are, and what percentage of the total tokens use a numerative form.

Table 10: Nouns that use a numerative with *trì*/3 and *ceithir*/4 combined, sorted by the number of numerative tokens from highest to lowest (all translations are from Bauer & Robertson (2016))

Noun	# of Numerative Tokens	Total Tokens	% of Numerative
bliadhna 'year'	163	168	97.02%
latha/là 'day'	92	153	60.13%
timcheall 'circuit [3]	30	30	100%
sgillinn 'pence, penny'	18	28	64.29%
oidhche 'night'	11	16	68.75%
fillte 'fold, ply'	9	11	81.82%
bunach 'squat, short'	8	9	88.89%
nì 'thing'	8	35	22.86%
cuairt 'time, cycle'	6	7	85.71%
tunna 'ton'	5	7	71.43%
uiread 'number, quantity'	5	5	100%
crann 'mast'	4	10	40%
cruinn 'round'	4	5	80%
seòrsa 'category, type'	4	16	25%
slat 'long stick, yard (measurement)'	4	5	80%
unnsa 'ounce'	4	8	50%
cas 'foot (body part), leg'	3	14	21.43%
ceàrn 'corner'	3	12	25%
duilleag 'leaf, page, sheet'	3	5	60%
eang 'foot, leg, hoof'	3	3	100%
tastan 'shilling'	3	19	15.79%
àird 'height/level'	2	4	50%
àirde 'point of the compass'	2	3	66.67%
bonn 'base, bottom'	2	8	25%
buille 'beat, hit, blow [noun]'	2	5	40%
caol 'sea channel'	2	6	33.33%
cùirt 'court'	2	2	100%
dual 'curl, fold'	2	3	66.67%
each 'horse'	2	10	20%
long 'ship'	2	2	100%
mathan 'bear'	2	3	66.67%
minidh 'awl'	2	2	100%
pearsa 'person'	2	5	40%
punnd 'pound (money)'	2	19	10.53%
rìgh 'king'	2	7	28.57%
seachdain 'week'	2	53	3.77%
troigh 'foot (measurement)'	2	28	7.14%
bann 'band, sash'	1	2	50%
bàr 'bar (establishment)'	1	3	33.33%
bean-ghlùine 'mid-wife'[4]	1	2	50%

---

[3] *ceithir thimcheall/thimchioll* is a fixed phrase. It means "all around" or "right around". All 30 tokens occur in this fixed phrase with *ceithir* 'four'.

biast 'beast, monster'	1	1	100%
bò 'cow'	1	1	100%
brag 'bang, crack (onomatopoeic)'	1	1	100%
braon 'drop of liquid'	1	1	100%
breab 'kick'	1	2	50%
buail 'step, degree'	1	1	100%
caochladh 'alteration, change'	1	1	100%
carbad 'carriage'	1	5	20%
cath 'battle, fight'	1	2	50%
cead 'right, approval'	1	1	100%
ceangail 'link, tie'	1	1	100%
ceathramh 'quarter'	1	6	16.67%
céile 'partner'	1	1	100%
céis 'case, envelope'	1	1	100%
ceum 'step'	1	2	50%
claiseach 'fluted blade'	1	1	100%
còrd 'cord'	1	1	100%
cothrom 'chance, opportunity'	1	2	50%
deuchainn 'exam, test'	1	1	100%
dòigh 'approach, method'	1	7	14.29%
dualach 'bush'	1	1	100%
duine 'person, man'	1	3	33.33%
dùthaich 'country, land'	1	1	100%
éireag 'cloudberry'	1	1	100%
fad 'length'	1	1	100%
fàth 'cause, reason'	1	1	100%
feasgar 'afternoon'	1	1	100%
fiacaill 'tooth'	1	7	14.29%
Gàidheal 'Gael'	1	1	100%
gaoid 'blemish'	1	1	100%
grod 'groove'	1	1	100%
ìobairt 'sacrifice, offering'	1	1	100%
ìochdar 'bottom, depth'	1	1	100%
iomlaid '(act of) changing money'	1	1	100%
leud 'width'	1	1	100%
lòn 'food, provisions'	1	2	50%
màlaid 'suitcase'	1	1	100%
mart 'cow'	1	11	9.09%
màs 'bottom, behind'	1	1	100%
meud 'size, amount'	1	1	100%
meur 'finger'	1	8	12.50%
mìos 'month'	1	61	1.64%
neach 'person'	1	1	100%
nochdadh '(act of) appearing'	1	1	100%
pinnt 'pint'	1	2	50%
ràith 'season'	1	5	20%

---

[4] The simple form *bean* 'woman, wife' is counted as one of the tokens, but only *bean-ghlùine* 'mid-wife' has a numerative in my data.

ràmh 'oar, paddle'	1	2	50%
rèidh 'flat, level'	1	1	100%
réite 'settlement, agreement'	1	1	100%
roinn 'department, section'	1	18	5.56%
ròlaist 'exaggeration, fib'	1	1	100%
roth 'wheel'	1	1	100%
saoghal 'world'	1	4	25%
sàth 'push'	1	1	100%
sgailc 'blow, knock'	1	1	100%
siola 'horse collar'	1	1	100%
slacan 'mallet'	1	1	100%
stiall 'streak, stripe'	1	1	100%
stuadh 'arch'	1	1	100%
suim 'sum'	1	1	100%
teist 'reputation, testimony'	1	1	100%
tiotan 'instant, jiffy'	1	1	100%
tùbh 'side, flank'	1	1	100%
turas 'journey, trip'	1	12	8.33%
uachdaran 'laird'	1	1	100%

From Table 10 we can see 105 nouns that were found using a numerative identical to a singular after a numeral. Of those, 68 were nouns that only used a numerative once. There seems to be semantic connections between which nouns appear in this table, such as measurement and money words. I go further in detail with this in Table 20, in §5.4. As with the duals in Table 5, this table demonstrates a wide range of frequencies of use for the numerative forms of these nouns. *Ceathramh* 'quarter', a numerative noted by Ó Maolalaigh (2013) only used a numerative for 16.67% of the time in my data and on the other side of the spectrum is *bliadhna* 'year' with a high number of numerative tokens making up 97.02% of the total tokens. This table highlights that the frequency of numerative use varies depending on the noun.

Table 11: Number of Numerative Tokens for *trì*, '3', *ceithir*, and '4' Combined by Date of Language in the Standard Category

Date of Language	# of Numerative Tokens	# of Total Tokens	% of Numeratives
1800-1849	74	331	22.36%
1850-1899	51	182	28.02%
1900-1949	114	477	23.90%
1950-1999	228	789	28.90%
2000-2016	24	80	30%

In Table 11, I present the use of numeratives diachronically. The time period with the highest percentage of use of numeratives is the 2000-2016 period, using a numerative for 30% of the

tokens. The group with the second highest use of numeratives is 1950-1999 with 28.90%, followed by 1850-1899 at 28.02%. The time period with the lowest percentage of use of numeratives is 1800-1849 at 22.36%, but 1900-1949 is very close at 23.90%. The difference between the highest percentage (30% for 2000-2016) and lowest percentage (22.36% for 1800-1849) is not statistically significant (Fisher's p = 0.1876). The difference between 1800-1849 and 1850-1899 is also not statistically significant (Chi-squared χ2=2.0453, df=1, p = 0.152681 and Fisher's p = 0.1631), as is the case with the difference between 1850-1899 vs 1900-1949 (Chi-squared χ2=1.1929, df=1, p = 0.274743 and Fisher's p = 0.3145); 1900-1949 vs 1950-1999 (Chi-squared χ2=3.7663, df=1, p = 0.052294 and Fisher's p = 0.0582); and 1950-1999 vs 2000-2016 (Chi-squared χ2=.0429, df=1, p = 0.835927 and Fisher's p = 0.8972). Based on this data, it appears that use of numeratives has stayed fairly consistent over time.

Table 12: Number of Numeratives with *trì* 'three', '3' *ceithir* 'four', and '4' Combined by Dialect Group in the Standard Category

Dialect Group	# of Numerative Tokens	# of Total Tokens	% of Numeratives
Canada	13	69	18.84%
Central	440	1632	26.96%
Peripheral	38	157	24.20%

The Central dialect group has the highest percentage of numeratives at 26.96%, followed by the Peripheral group at 24.20%, and finally the Canadian dialect group at 18.84%. The difference between the Canadian dialect and Central dialect is not statistically significant (Chi-square χ2=2.2341, df=1, p = 0.134999 and Fisher's exact test p = 0.164). The difference between the Central and Peripheral dialects is also not statistically significant (Chi-square χ2=0.556, df=1, p = 0.455883 and Fisher's exact test p = 0.5089). The difference between the Canadian and Peripheral dialects is also not statistically significant (Chi-square χ2=0.789, df=1, p = 0 .374388 and Fisher's exact test p = 0.4896). The differences between the dialects are not statistically significant and the data indicates that the dialect has no significant impact on the use of numeratives, especially since they all were similar in the percentage of numeratives.

In order to try to see a pattern in numerative use, I analyze each of the nouns with at least ten tokens using a numerative, separated by time period and dialect. The nouns that are analyzed individually are listed in Table 13, below.

Table 13: Nouns with More than Ten Tokens using a Numerative

Noun	# of Numeratives	# of Total Tokens	% of Numerative
bliadhna 'year'	163	168	97.02%
latha/là 'day'	92	153	60.13%
timcheall 'circuit'[3]	30	30	100%
sgillinn 'pence, penny'	18	28	64.29%
oidhche 'night'	11	16	68.75%

From the nouns in Table 13, *timcheall* 'circuit' will not be analyzed because it is in a fixed phrase with *ceithir* 'four', so all tokens were numeratives 100% of the time[3].

For Tables 14-17, the fractions indicate the number of numerative tokens over the total number of tokens for that noun for each time period and dialect. Additionally, the Canadian and Peripheral dialects have too few tokens, if any at all, so the only dialect that will be analyzed more in-depth is the Central dialect.

Table 14: Tokens of Numeratives for *bliadhna* 'year' by Dialect and Date of Language

Dialect/Date of Language	1800-1849	1850-1899	1900-1949	1950-1999	2000-2016
Canada	0	0	0	9/9 [100%]	0
Central	13/13 [100%]	22/25 [88%]	35/35 [100%]	66/67 [98.51%]	11/11 [100%]
Peripheral	3/3 [100%]	0	0	4/5 [80%]	0

For the Central dialect, *bliadhna* 'year' is used as a numerative for 100% of the tokens for 1800-1949, 1900-1949, and 2000-2016. A numerative is used for 88% of the tokens in 1850-1899 and for 98.51% of the tokens in 1950-1999. The differences between 1800-1849 vs 1850-1899 (Fisher's p = 0.5377), 1850-1899 vs 1900-1949 (Fisher's p = 0.0672), 1900-1949 vs 1950-1999 (Fisher's p = 1), and 1950-1999 vs 2000-2016 (Fisher's p = 1) are all not statistically significant. This means that from this data, it appears that numerative use of *bliadhna* 'year' in the Central dialect has not changed significantly over time.

Table 15: Tokens of Numeratives for *latha/là* 'day' by Dialect and Date of Language

Dialect/Date of Language	1800-1849	1850-1899	1900-1949	1950-1999	2000-2016
Canada	0	0	0	1/5 [20%]	0
Central	12/29 [41.38%]	6/17 [35.29%]	22/40 [55%]	39/48 [81.25%]	8/8 [100%]
Peripheral	1/3 [33.33%]	0	0	3/3 [100%]	0

In the Central dialect for *latha/là* 'day', the percentage of numeratives decreases from 41.38% in 1800-1849 to 35.29% in 1850-1899. This decrease is not statistically significant

(Chi-square χ2=0.1666, df=1, p = 0 .683135 and Fisher's exact test p = 0.7613). After this, the numerative use of *latha/là* 'day' increases in every subsequent time period. The difference between 1850-1899 (which used numeratives for 35.29% of the tokens) vs 1900-1949 (55% of the tokens used a numerative) was also not statistically significant (Chi-square χ2=1.8536, df=1, p = 0.173364 and Fisher's exact test p = 0.2483). The increase in the percentage between 1900-1949 and 1950-1999 from 55% to 81.25%, is statistically significant (Chi-square χ2=7.0689, df=1, p = 0 .007844 and Fisher's exact test p = 0.0107). The difference for 1950-1999 (uses a numerative for 81.25% of the tokens) vs 2000-2016 (uses a numerative for 100% of the tokens) was not statistically significant (Fisher's p = 0.3294). The difference between 1850-1899 vs 2000-2016 was statistically significant (Fisher's p = 0.0029). Additionally, the difference between 1850-1899 and 1950-1999 is also statistically significant (Chi-square χ2=12.4464, df=1, p = 0 .000419 and Fisher's exact test p = 0.0014). This means that the date of language does have an impact on the use of the numerative form of *latha/là* 'day' and the increase in the use of the numerative form is significant.

Table 16: Tokens of Numeratives for *sgillinn* 'pence, penny' by Dialect and Date of Language

Dialect/Date of Language	1800-1849	1850-1899	1900-1949	1950-1999	2000-2016
Canada	0	0	0	0/3	0
Central	5/6 [83.33%]	1/2 [50%]	3/7 [42.86%]	8/8 [100%]	0
Peripheral	0	0	0	1/2 [50%]	0

The use of *sgillinn* 'pence, penny' in the Central dialect has an overall decrease in the amount of numerative tokens out of the total number of tokens. The differences that are not statistically significant are: 1800-1849 vs 1850-1899 (Fisher's p = 0.4643), 1850-1899 vs 1900-1949 (Fisher's p = 1), 1950-1999 vs 2000-2016 (Fisher's p = 1), 1800-1849 vs 1900-1949 (Fisher's p = 0.2657), 1800-1849 vs 1950-1999 (Fisher's p = 0.4286), and 1850-1899 vs 1950-1999 (Fisher's p = 0.2). The difference that was statistically significant is 1900-1949 vs 1950-1999 (Fisher's p = 0.0256). Because of almost all of the differences between the time periods not being statistically significant, there appears to be no significant diachronic change in numerative use for the Central dialect for *sgillinn* 'pence, penny' except for the increase between 1900-1949 and 1950-1999. The increase in use between these two time periods could truly reflect an increase in numerative use, but it could also be due to representation of different text genres in the corpus.

Table 17: Tokens of Numeratives for *oidhche* 'night' by Dialect and Date of Language

Dialect/Date of Language	1800-1849	1850-1899	1900-1949	1950-1999	2000-2016
Canada	0	0	0	0	0
Central	4/4 [100%]	0/1	1/4 [25%]	4/6 [66.67%]	2/2 [100%]
Peripheral	0	0	0	0	0

The numerative use of *oidhche* 'night' decreased at first from 1800-1849 to 1850-1899 and then increased again. However, the decrease in numerative use from 1800-1849 and 1850-1899 was not statistically significant (Fisher's $p = 0.2$), and this also goes for the increase between 1850-1899 and 1900-1949 (Fisher's $p = 1$). The differences between 1900-1949 and 1950-1999 (Fisher's $p = 0.5238$), and 1950-1999 and 2000-2016 (Fisher's $p = 1$) were also not statistically significant. Based on this data, it appears that the use of the numerative form of *oidhche* 'night' has not changed significantly over time.

5   Discussion

5.1   Dual

As shown in Table 5, in the data there are 47 nouns that are found using a dual form after the numeral two. Of those 47 nouns, 25 only use the dual form once. Only three nouns have more than ten tokens of the dual: *làmh* 'hand', *cluas* 'ear', and *cas* 'foot (body part), leg'. With the nouns that only had one dual token, it is hard to know from this data whether it is a universal dual form or if only one author uses it. One way to determine this could be another possible research project, which could remove the restrictions that this study had (date of language and geographic origin) and search the corpus for the nouns with only one dual token in my results and see if the dual form shows up more than once.

Another possibility for the nouns with only one dual token is that it could be just an orthographic error. This is also a problem for nouns with a small amount of dual tokens that all come from the same text. For example *craiceann* 'skin, parchment' has three tokens using a dual form, but they all come from the same text (CnG: 49). This text is not particularly heavy in the amount of dual tokens it contained in my data and only contributes one other noun with a single dual token, which is *boiceann* 'goatskin'. It could be that *craiceann* 'skin, parchment' with all three dual tokens from one text does have a universal dual form and just did not occur in my data with more tokens or it could be that the author was overgeneralizing. It could be argued that the authors using new dual forms by analogy is insight into what kind

of words Scottish Gaelic speakers generalize as having dual forms. This could happen if there is a semantic similarity between many nouns that use a dual, such as parts of the body, and then the author encounters a noun which is similar to another body part and uses a dual form when others had not used a dual form for that particular noun in other occurrences. This is one possibility for dual forms with only one token. It should also be noted that the nouns with only one dual form token in my data do not come from a single text; it is not just one author using all of these single dual tokens. Regardless, it is interesting to observe which nouns have dual forms. Many of the nouns denote parts of the body, like 'thumb', 'tooth', 'knee', and 'hand'. In Table 18 (below), I sort the nouns with dual forms into semantic categories.

Table 18: Semantic Categories of Nouns that have a Dual Form

Semantic Category	Nouns
Natural Pairs	*làmh* 'hand', *cluas* 'ear', *cas* 'foot (body part), leg', *bas* 'palm (of a hand)', *slinnean* 'shoulder blade', *glùn* 'knee', *bròg* 'shoe', *òrdag* 'thumb', *lurgann* 'shaft, shin, leg'
Non-paired body parts	*craiceann* 'skin, parchment', *fiacal* 'tooth', *tòn* 'backside, bum', *aodann* 'face', *cliathach* 'side (of a being)', *corrag* 'finger, toe', *cliathach* 'side (of a being)'
People	*buidheann* 'band, faction, group', *sluagh* 'folk, people', *ceatharna* 'troop', *bean* 'woman, wife', *caileag* 'girl', *oifigear* 'officer', *teaghlach* 'family'
Measurement	*earrann* '(part of an) area, piece', *òirleach* 'inch (measure)'
Nautical	*eathar* 'small boat/vessel', *long* 'ship', *ràmh* 'oar, paddle'
Animal	*spàg* 'paw, limb (of an animal)', *boiceann* 'goatskin', *cearc* 'hen, chicken', *adhrac* 'horn'
Nature	*clach* 'rock, stone', *coilleag* 'shoot, sprout', *eilean* 'isle, island', *glac* 'glade', *tulach* 'hillock', *craobh* 'tree'
Miscellaneous	*cànan* 'language', *focal* 'word', *dìnnear* 'dinner', *faobhar* 'edge (of a blade)', *rathad* 'road', *teud* 'cord, string', *searrag* 'flask, bottle', *basgaid* 'basket'

In Table 18, it is easy to see that the Natural Pairs category has the most nouns with a dual form in my data. This was the category I was expecting to be well-represented by duals since all of the duals in Irish denote natural pairs. There are also many body parts that do not come in pairs that have distinct dual forms. Perhaps these are used frequently enough with 'two' that they retain a distinct dual form. It is interesting that *spàg* 'paw, limb (of an animal) appeared in a dual form because animals have four legs or paws, compared to *cas* 'foot (body part), leg' which is a natural pair. The People category also contains seven nouns that have a distinct dual form. This is unsurprising because nouns denoting animates are more likely to retain an irregular form, as mentioned in §2.5 (Corbett 2009: 6, 55). Other than the natural

pairs, there does not seem to be a strong overall trend with the semantic categories of the nouns that have a dual form.

When looking at the use of dual diachronically (Table 6), the percentage of use stayed fairly consistent until 1900-1949, where it peaked at 8.12% and then decreased over time, ending up at 1.76% in 2000-2016. The increase in use of dual in 1900-1949 is interesting, especially since it is statistically significant and the time periods before and after it are at about the same percentage of dual use. This increase does not necessarily mean that dual was suddenly used much more in 1900-1949 before dropping back off again; it could instead be because of the texts in the corpus. It could be that the texts for 1900-1949 are comprised mainly of more conservative texts, such as religious texts and poetry. For a future research project, the texts of the corpus could be divided into genres to see how many texts of each genre each time period contains.

According to my data, the use of dual is decreasing after the increase in 1900-1949. As mentioned above, the decrease from 1900-1949 to 1950-1999 was not significant, but the decrease from 1900-1949 to 2000-2016 is significant. With time periods that are further apart it is easier to see a change since it has had more time to develop. It could be that the difference between 1900-1949 and 1950-1999 is not enough time to be able to see a significant decrease in numerative use. Increasing the difference in time also indicates a clearer, and significant, decrease in dual usage. As discussed in §2.4 and §2.5, it is not uncommon in general for languages to lose the dual as a full number category or for dual to be retained only by a small group of nouns in a language, as is the case with Modern Hebrew, Maltese, Welsh, and Irish. As seen in these languages, the nouns that tend to keep their dual form the longest are nouns that are often associated with natural pairs (such as hands, ears, feet) and nouns used for counting or measuring (such as 'day' see example (1)-(3)) because of the frequency of use. I also found this in my results, with natural pairs being the semantic category with the most dual forms. Nouns used for counting or measuring and especially natural pairs are not just frequent in general, but are also very frequent with 'two', so this is likely why their numerative forms persist.

I also looked at the use of dual across dialects to see if there is a difference between them (Table 7). This ended up being more difficult than I expected due to the small amount of tokens for the Canadian dialect (seven total tokens). Because of the token amount, it cannot be determined if the dual has disappeared from the Canadian dialect or if the data just does not contain any dual tokens. The Peripheral dialect was more represented than the Canadian dialect but still had many fewer tokens than the Central dialect, but the difference between

the Peripheral and Central dialects is not significant. Therefore, the data appears to show that there is no significant difference in dual usage across dialects. I do think that a separate research project focusing on the Peripheral and Canadian dialects individually could be useful, not just for numerative use but for overall knowledge about language patterns in these two dialects. This would require different corpora, assembled with texts from these dialects specifically, but these corpora would add great momentum to study the Peripheral and Canadian dialects of Scottish Gaelic.

In order to try to see a clearer pattern in numerative use, both across time and dialects, I took the nouns with the highest number of tokens using dual and analyzed them. Specifically, I target nouns with more than ten dual tokens, which was three nouns. The two nouns that were analyzed individually were *làmh* 'hand' and *cas* 'foot (body part), leg'. *Cluas* 'ear' had the second-highest number of tokens using dual, but all of the tokens that I found used dual, so looking at its use would not have shown any change in use since it always used dual in my data. *Làmh* 'hand' had a total of 76 tokens, 58 of which used the dual form (76.32%). *Cas* 'foot (body part), leg' had a total of 26 tokens and of those, 16 used the dual form (61.54%). Both nouns have no tokens for the Canadian dialect and the Peripheral dialect has very few tokens overall, with three tokens for *làmh* 'hand' and two for *cas* 'foot, leg'. The Central dialect was the only one of the three that had enough tokens to see a pattern across the time periods, so that is the dialect I will be focusing on for these two nouns.

While looking at how the use of the dual of *làmh* 'hand' is used over time in the Central dialect, there are some time periods that have differences that are statistically significant. The differences between 1800-1849 vs 1850-1899 and 1850-1899 vs 1900-1949 are both statistically significant differences, but the difference between 1800-1849 and 1900-1949 was not. This means that the date of language does have some impact on the use of the dual form of *làmh* 'hand', but only for the decrease from 1800-1849 (87.5%) to 1850-1899 (0%) and then the increase in 1900-1949 (80.77%). The increase between 1850-1899 and 1900-1949 could reflect an actual increase in use of the dual form of *làmh* 'hand', but it also could be due to many other factors, such as genre of the texts, as mentioned earlier, or just a higher representation of *làmh* 'hand' in the texts.

Cas 'foot (body part), leg' has fewer tokens than *làmh* 'hand' and the differences between all of the time periods were not statistically significant. Due to the small amount of tokens and the lack of significance, it is hard to make any conclusions, but it appears that dual use of *cas* 'foot (body part), leg' has stayed fairly consistent across time and has not changed significantly.

My research questions were focused on what nouns had dual forms and whether there was dialectal and diachronic variation in their use. Based on my data, in the Central dialect there is an increase in the use of dual from 1850-1899 to 1900-1949 and then from the early 1900s, there has been a decrease in dual tokens, though this was not significant. Further studies could try to repeat this experiment, but limit the results to specific genres of text to try to capture more casual language use. In my study, since I used all genres, it is possible that a pattern was obscured by having many different registers. The Peripheral dialect and the Canadian dialect have too few tokens to know how prevalent duals are and how the frequency of their use has changed over time. The decrease in the amount of duals used over the 20th century into the 21st also coincides with my hypothesis that dual use has decreased over time. As discussed in §2.1, Scottish Gaelic is virtually no longer being acquired as an L1 in the home (Moseley 2010). This is an important thing to note because in §2.7 we learned that linguistic complexity decreases when L2 learners become a large portion of the speakers of a language (McWhorter 2008: 168, Karlsson et al. 2008: viii). Having a distinct dual form makes Scottish Gaelic more complex, so it is unsurprising that in my data the use of the dual has decreased from 1900-1949 to 2000-2016.

5.2   Numeratives Identical in Form to Singular

In the data, there are 105 nouns that are found using a numerative identical to a singular after a numeral. Of those, 68 are nouns that only use a numerative once. As with the dual tokens that only occur once in the data, these nouns are also from various different texts and not just from one author who uses many numeratives. For the nouns that have only one token with that token being a numerative, it is hard to know if these nouns tend to use a numerative and it is just not a very common word or if the author uses a numerative because of analogy. It could be argued that even accidentally using a numerative could give some insight into how the noun is viewed. For instance if a word is a measurement word, like 'ounce', the speaker may automatically use a numerative identical to singular in form because they categorize it with other measurement words, like 'ton', which did use a numerative form 71.43% of the time in my corpus data.

In Table 11, I present the use of numeratives diachronically. The difference between the highest percentage (30% for 2000-2016) and lowest percentage (22.36% for 1800-1849) is not statistically significant. This data indicates that there was an increase in 1850-1899, a decrease in 1900-1949 and then an increase through the next two 50-year periods, ending up

with an overall increase when comparing the 1800-1849 time period to the 2000-2016 time period. However, the data available did not result in statistical significance. This could indicate that the noun itself is more important for numerative use than the time period.

The percentages of numeratives out of the total tokens in the Standard category across dialectal groups were also fairly consistent (Table 12). This would also seem to suggest that dialect does not impact the use of numeratives either. The Central dialect has many more tokens overall in the corpus, at 1,632 for the searches for three and four combined, while the Peripheral dialect had 157, and the Canadian dialect had only 69 total tokens. Because the Canadian and Peripheral dialects had so many fewer tokens than Central, it is hard to say if the lower frequency of numerative use accurately reflects the language used or if it is just because of having less data in the corpus. The difference between the Central dialect and the Peripheral dialect, the Central and the Canadian dialect, and the Canadian and Peripheral dialects were all not statistically significant. Based on the data, it appears that dialect does not significantly impact the use of numerative.

There were five nouns that had more than ten tokens using a numerative. These five nouns are: *bliadhna* 'year', *latha/là* 'day', *timcheall* 'circuit', *sgillinn* 'pence, penny', and *oidhche* 'night'. All of the tokens for *timcheall* 'circuit' were numeratives, so it is not analyzed because it cannot be used to look at a pattern of change since it is always in a fixed phrase. All four of the analyzed nouns lack enough tokens in the Peripheral and Canadian dialects to observe a change in use. While it is possible that the numerative forms have disappeared from these dialects, based on my data it is impossible to know because the non-numerative number of tokens is also very small. A possible future research project could be to remove the date of language restriction and look at all Canadian and Peripheral dialect texts to see whether there is a change in numerative use.

In Table 14, we can also see that overall, *bliadhna* 'year' has a high percentage of numerative use, with the lowest percentage being 88% for 1850-1899 and three of the time periods having 100% of the tokens appearing as numeratives. From the data, it looks like numerative use has stayed fairly consistent diachronically in the Central dialect. None of the differences between the time periods were statistically significant, which could indicate that numerative use of *bliadhna* 'year' does not change significantly over time because it is consistently being used as a numerative. In addition to this, *bliadhna* 'year' also uses a numerative at a very high frequency. Out of all the nouns with a numerative identical in form to singular, *bliadhna* 'year' used a numerative for 97.0% of the tokens in my data (Table 10). Instead of the date of language having a correlation with the amount of numerative use,

numerative use could be based on the noun itself. The Canadian dialect had only nine tokens and they were all in 1950-1999, so my data cannot be used to say anything about how the use of numeratives has changed, but as stated above, a corpus study focused on Canadian texts could be a future research project. Similarly, the Peripheral dialect had six total tokens, three in 1800-1849 and three in 1950-1999. The data does show an increase in use of numeratives (33.33% in 1800-1949 and 100% in 1950-1999), but due to the small number of tokens and because the difference between those two time periods is not statistically significant, it could be that numerative use is staying consistent or it could be that I do not have enough data to see a clear pattern.

With *latha/là* 'day' in Table 15, we once again do not have enough tokens for the Canadian and Peripheral dialects to be able to observe if numerative use is changing or not. This does not necessarily mean that *latha/là* 'day' is not used as a numerative in these dialects, it could just be that there are not many tokens for *latha/là* 'day' in the corpus. The number of tokens could also be low due to the restrictions of this project, which is that texts need a date of language and geographic origin. According to my data, for the Central dialect, there is a gradual increase in the amount of numerative forms used for *latha/là* 'day'. The difference between 1900-1949 and 1950-1999, 1850-1899 and 2000-2016, and 1850-1899 and 1950-1999 were all statistically significant, so I can determine that as the date of language progresses, there is an increase in the numerative use of *latha/là* 'day'.

For *sgillinn* 'pence, penny', the Central dialect can be seen in Table 16 as having a decrease from 1800-1850 (83.33%) all the way to 1900-1949 (42.86%) and then an increase in 1950-1999 to all eight tokens using a numerative. There are no tokens in the 2000-2016 time period, in any form, but that does not mean that *sgillinn* 'pence, penny' is no longer being used, it is more likely that it just did not show up in the data due to the restrictions or having fewer texts from 2000-2016 in the corpus. The difference between 1900-1949 and 1950-1999 is statistically significant, but all the other differences between the time periods were not statistically significant. The large amount of not significant differences could mean that for *sgillinn* 'pence, penny' there is no diachronic changes in numerative use based on my data. The increase between 1900-1949 and 1950-1999 is statistically significant, which may reflect an actual increase in use or it may reflect the (possibly conservative) register used in the texts from that time period.

The last noun to be analyzed individually is *oidhche* 'night' in Table 17. When distributed across time periods, the total token amounts are quite small, for example 1850-1899 with only one token. Because the token amount is so small, it is more difficult to see a change in

use, but it does seem to overall stay fairly consistent, with some decrease in use in both 1850-1899 and 1900-1949 especially. However, none of the significance tests I did on the differences between the time periods are statistically significant, so based on my data, it appears that the date of language does not have a significant impact on the use of numerative for *oidhche* 'night'.

There was a small number of nouns that used both a dual form and a numerative identical in form to a singular in my results. These nouns were: *bean* 'woman, wife', *cas* 'foot (body part), leg', *fiacal* 'tooth', *long* 'ship', and *ràmh* 'oar, paddle'. This is an interesting group of animate nouns, inanimate nouns, and nouns for parts of the body. This would seem to indicate that these words are used frequently enough with two as well as with numerals three and above to have two separate numerative forms that have persisted diachronically. In both categories of numeratives, there were many human nouns, body parts, and nautical terms, so this list is not completely unsurprising, but I would have expected there to be some measurement or counting terms as well. If future research expands on numeratives in Scottish Gaelic, it would be interesting to see if the list of nouns with more than one numerative form grows as well and what semantic categories such nouns belong to.

Based on my data, the use of numeratives in the Central dialect seems to be increasing overall starting in 1900-1949. For each noun that was analyzed individually, there was a decrease in numerative use in 1850-1899. I believe that this is due to the number of texts since this time period was the 50-year group with the smallest number of texts. The 2000-2016 group had the smallest number of texts overall, but it also represented a smaller amount of time. The 1850-1899 time period had almost 200 total tokens less than the next smallest amount and about 700 less than the largest. For the results of the four nouns with the most numerative tokens, *bliadhna* 'year' appeared to be used as a numerative fairly consistently and at a high percentage of the tokens. After the decrease in 1850-1899, *là/latha* 'day', *sgillinn* 'penny, pence', and *oidhche* 'night' all appeared to increase in numerative use through the 2000-2016 time period. For *là/latha* 'day' and *sgillinn* 'penny, pence', the increase was statistically significant, which means that numerative use does increase as the date of language progresses. For dialectal differences, due to the small amount of tokens, no change in the use of numeratives could be observed in the Peripheral and Canadian dialects. The small amount of tokens could be due to numerative forms being lost in these dialects, but because the total number of tokens was also small, I would guess that these two dialects have fewer tokens because of having less material in the corpus.

For my results, my hypothesis was that nouns used frequently with numerals would retain their numerative forms since L2 learners could hear them frequently enough to learn them as phrases instead of as exceptions to a rule. I also hypothesized that numerative use would otherwise decrease, because of Scottish Gaelic losing complexity due to a large amount of L2 learners and the nouns levelling to more typical forms (see §2.7). However, based on the data, it appears that my hypothesis is incorrect since most of the numeratives increased in use overall and with three of the four nouns that used numerative most often. This could also be supported by the fact that the lexemes which have numeratives frequently in numeral +noun phrases (such as *bliadhna* 'year') stay fairly stable over time. Many of the numeratives were nouns for time, measuring, and money, which occur frequently with numbers and it seems that they occur frequently enough to be used as numeratives in either a stable way (all of the numeratives combined, *bliadhna* 'year', and *oidhche* 'night') or an increasing way (*latha/là* 'day' and *sgillinn* 'penny, pency').

5.3   Comparison to Previous Work

Not including the nouns for numerals, this project has 20 nouns that use a numerative identical in form to a singular that were the same nouns that Ó Maolalaigh (2013: 132) found in his research, which leaves 20 nouns that he found that I did not. There are many reasons that there are numeratives that did not overlap: Ó Maolalaigh (2013) was looking at all numbers three through ten, while I only looked at three and four; he also used texts for his data that lacked the necessary geographic origin and date of language that I wanted for analysis, so he had more texts available to use in the 20th century than I did; and seven of the nouns that Ó Maolalaigh (2013) counted as having numeratives are nouns that I put into the Complex Number category, such as *fichead* 'twenty', *mile* 'thousand', and *mìleamh* 'thousandth'. I did not count these as nouns that use a numerative because in typical numeral phrases, they use a singular form (see example (16) in §3.3 or §2.6 for a complete discussion on number and numeratives in Scottish Gaelic). The nouns that were in common between this project and his were: *ceàrn* 'quarter', *ceathramh* 'stanza' (lit. 'quarter'), *bliadhna* 'year', *cuairt* 'time, cycle', *duine* 'person, man', *latha/là* 'day', *oidhche* 'night', *seachdain* 'week', *fad* 'length', *slat* 'yard (measurement)', *meud* 'size', *uiread* 'quantity, amount', *tunna/tonna* 'ton', *punnd* 'pound (currency)', *sgillinn/sgilling* 'penny', *tastan/tasdan* 'shilling', *troigh* 'foot (measurement)', *pearsa* 'person', *seòrsa* 'type', and *ùnnsa* 'ounce' (Ó Maolalaigh 2013: 132). As Ó Maolalaigh (2013) put the numeratives he found into semantic categories (see

Table 2, in §2.6), I also sorted the numeratives I found into semantic categories in the table below:

Table 19: Semantic Categories of Nouns that use a Numerative Identical in Form to a Singular

Semantic Category	Nouns
Time	*bliadhna* 'year', *latha/là* 'day', *oidhche* 'night', *seachdain* 'week', *feasgar* 'afternoon', *mìos* 'month', *ràith* 'season', *tiotan* 'instant, jiffy'
Animals	*biast* 'beast, monster', *bó* 'cow', *each* 'horse', *mart* 'cow', *mathan* 'bear',
Of the Body	*cas* 'foot, leg', *eang* 'foot, leg, hoof', *fiacal* 'tooth', *gaoid* 'blemish', *màs* 'bottom, behind', *meur* 'finger', *tùbh* 'side, flank', *dual* 'curl, lock of hair'
Measurements	*ceathramh* 'quarter', *cuairt* 'time, cycle', *meud* 'size, amount', *slat* 'long stick, yard (measurement)', *troigh* 'foot (measurement)', *tunna* 'ton', *uiread* 'number, quantity', *unnsa* 'ounce', *àird* 'height/level', *braon* 'drop of liquid', *buail* 'step, degree', *ceum* 'step', *cruinn* 'round', *fad* 'length', *ìochdar* 'bottom, depth', *leud* 'width', *pinnt* 'pint', *timcheall* 'circuit', *bunach* 'squat, short', *suim* 'sum'
Money	*punnd* 'pound (money)', *sgillinn* 'pence, penny', *tastan* 'shilling', *iomlaid* 'exchange, barter'
People	*duine* 'person, man', *pearsa* 'person', *bean-ghlùine* 'mid-wife', *céile* 'partner', *Gàidheal* 'Gael', *neach* 'person', *rìgh* 'king', *uachdaran* 'laird'
Nautical	*àirde* 'point of the compass', *caol* 'sea channel', *crann* 'mast', *long* 'ship', *ràmh* 'oar'
Nature	*dualach* 'bush', *duilleag* 'leaf, page, sheet', *éireag* 'cloudberry' *saoghal* 'world'
Place	*dùthaich* 'country, land', *bàr* 'bar (establishment)', *bonn* 'base, bottom', *cùirt* 'court', *roinn* 'department, section'
Shapes	*ceàrn* 'corner', *fillte* 'fold, ply', *grod* 'groove', *rèidh* 'flat, level', *stiall* 'streak, stripe', *stuadh* 'arch'
Objects	*ceangail* 'link, tie', *céis* 'case, envelope', *claiseach* 'fluted blade', *còrd* 'cord', *màlaid* 'suitcase', *minidh* 'awl', *nì* 'thing', *roth* 'wheel', *slacan* 'mallet', *bann* 'band', *carbad* 'carriage', *siola* 'horse collar'
Actions	*breab* 'kick', *buille* 'beat, hit, blow', *nochdadh* '(act of) appearing', *sàth* 'push', *sgailc* 'blow, knock', *caochladh* 'alteration, change', *ìobairt* 'sacrifice, offering', *ròlaist* 'exaggeration, fib', *brag* 'bang, crack (onomatopoeic)'
Abstract	*cead* 'right, approval', *cothrom* 'chance, opportunity', *fàth* 'cause, reason', *réite* 'settlement, agreement', *teist* 'reputation, testimony', *deuchainn* 'exam, test', *seòrsa* 'category, type', *dòigh* 'approach, method'
Miscellaneous	*cath* 'battle, fight', *lòn* 'food, provisions', *turas* 'journey, trip'

The semantic categories I have in common with Ó Maolalaigh (2013) are: time, length/distance (which I put into "measurements"), measure, weight (also put into

"measurements"), currency, and people. The range of nouns with numeratives that I found was much broader than Ó Maolalaigh (2013), with animals, items, shapes, etc. I am not surprised by this because Ó Maolalaigh (2013) limited his texts to the 20th century, so he had a much smaller timeframe than this project.

Because the grammar of Irish is slightly different from Scottish Gaelic, I will only be focusing on comparing the nouns themselves that use numeratives, as opposed to what form of numerative they use (i.e. identical in form to a singular or a unique form). The lists of numeratives in Irish are not exhaustive, as mentioned in §2.5, but I will compare the numeratives I found in my results with the known numeratives in Irish. The numeratives across Irish dialects that I also found in my data are: *orlach* 'inch', *troigh* 'foot (measure)', *slat* 'rod (measure), yard', *bliain* 'year', *seachtain* 'week', *scilling* 'shilling' (Ó hAnluain 1999: 70), *cos* 'foot', *bos* 'palm (of the hand), *bróg* 'shoe', cluas 'ear', and *lámh* 'hand' (Christian Brothers 1960: 82). From Munster Irish: *punt* 'pound', *turas* 'journey' (Ó Siadhail 1982: 102-104, 1989: 166-168), *bó* 'cow', *méar* 'finger', *focal* 'word', and *piunt* 'pint' (Ó Sé 2000: 225-228). 'Horse' and 'man' were also nouns with numeratives that Munster Irish and Scottish Gaelic had in common, but the words were different (*fear* 'man' in Scottish Gaelic did not appear as a numerative, but *duine* 'man' did and *each* 'horse' occurred in Scottish Gaelic but it does not correspond etymologically to *capall* 'horse' in Irish). For Connacht Irish the numeratives that are also found in Scottish Gaelic are: *ceathrú* 'quarter', *cloch* 'stone (of weight)', *dual* 'strand (of hair)' *lámh* 'hand', and *punt* 'pound' (Ó Siadhail 1982: 102-104, 1989: 167-168). For Ulster Irish the overlapping numeratives with my results are: *cos* 'foot', *cloch* 'stone (of weight)', *méar* 'finger', and *bascáid* 'basket(ful)' (Ó Baoill 1999: 104-105). The Irish dialect with the most numeratives in common with Scottish Gaelic is Munster with six numeratives. Connacht has five numeratives in common with Scottish Gaelic and Ulster has four, so all three dialects were fairly similar in overlap with Scottish Gaelic. There were six numeratives that are general to all dialects of Irish that were also found in Scottish Gaelic.

The nouns that have numeratives in Irish but not in my data were *pingin* 'penny' (which Ó Maolalaigh (2013) did have listed as a numerative in his data) and *uair* 'time, occasion' (also a result Ó Maolalaigh (2013) had) for all dialects of Irish (Ó hAnluain 1999: 70). For Munster Irish, the numeratives I did not find in my data were *fear* 'man', *real* 'sixpence', *nóimint* 'minute' (a result Ó Maolalaigh (2013) had), *bád* 'boat(ful) (Ó Siadhail 1982: 102-104, 1989: 166-168), *cnámh* 'bone', *maidean* 'morning', *naomhóg* 'coracle, small boat', *capall* 'horse', *fód* 'strip of land', *soitheach* 'vessel, jar', *tigh* 'house', *bua* 'life, gift',

suíochán 'seat', *gort* 'field' (Ó Sé 2000: 225-228). For Connacht, the numeratives I did not find in my data were *bord* 'load', *galún* 'gallon', *mála* 'bag(ful)', *stór* 'storey', and *ualach* 'load' (Ó Siadhail 1982: 102-104, 1989: 167-168). The numeratives from Ulster Irish that I did not find in my results are: *capall* 'horse', *fód* 'strip of land', *ceann* 'head (as a unit)', *doras* 'door', *uan* 'lamb', and *teanga* 'language' (Ó Baoill 1999: 104-105). While there were many numeratives in Scottish Gaelic and Irish that did not overlap, semantically the Irish numeratives are very close to my results.

The known numeratives for Irish and the numeratives I found in my data are likely not all the numeratives that exist in these languages since no complete corpus study of numeratives has been done for either. However, the numeratives I found for Scottish Gaelic have a large amount of overlap with its numeratives in Irish, with Munster Irish having the most numeratives in common with Scottish Gaelic. It is not surprising to have this overlap in numeratives because of their common origin, but it is interesting that the same numeratives have persisted in Irish and Scottish Gaelic.

5.4   Summary of Results

Based on my data, there is a slight decrease in the amount of duals being used over time, at least in the Central dialect. The differences between dialects are not statistically significant, so based on this data, there does not appear to be variation in dual usage based on the dialect. The Peripheral dialect and the Canadian dialect have too few tokens to know how prevalent duals are, if dual forms still exist in these dialects, and whether the frequency of their use has changed over time. My data also indicates that the use of numeratives that are identical in form to singulars has stayed consistent over time. The differences between dialects for these numeratives are also not statistically significant. However, for both duals and numeratives identical in form to singular, there are differences between time periods that are statistically significant, which indicates that the date of language does have an impact on use of these forms. There could also be other variables that are impacting numerative use, such as genre. If the corpus had a large amount of more conservative texts, such as religious texts, in one 50-year time period, it could skew the numerative use to be higher than what everyday language uses.

In terms of complexity, it makes sense why use of the dual would decrease because a dual is an additional form to remember and increases complexity and the amount of L2 learners who make up the Scottish Gaelic users is larger now. It may be surprising that the

use of numeratives identical in form to singular has increased since 1900-1949 because these nouns could be considered as exceptions to the typical pattern, which would increase the complexity. However, it could be that nouns with numerative forms are frequent enough in use that L2 learners are able to learn them. According to my data, even with a large amount of the speakers of Scottish Gaelic being L2 learners, there has been no simplification for numeratives identical in form to singular.

Perhaps if this research project is repeated or replicated, the researcher could take a random sampling of a certain amount for each dialect group. This way, the sample size would be the same and then we could see how many numeratives are used within that sample and compare across the dialects. A fixed sample size could also help with time periods that were not as well represented in the number of total tokens.

Due to the fact that Scottish Gaelic did not have a standardized orthography for most of the time when the texts from the corpus were published, it is possible that spellings for numerals have been missed.

6   Conclusion

The questions this research aimed to answer were:

(1) What nouns use a numerative identical in form to singular with numbers three and four?
(2) Which nouns have dual forms?
(3) How consistently do the nouns that have a numerative form use the numerative form?
(4) Does the use of numeratives vary between different dialects?
(5) How does noun form use vary diachronically? Is the use of numeratives changing over time?

In my data, I found 47 nouns that use a dual form and 105 nouns that use a numerative identical in form to the singular form. Nouns that tended to use dual were natural pairs, like 'hand' and 'ear', while singular-looking numeratives were strongly associated with nouns that denote time, like 'day' and 'year', money terms such as 'penny, pence' and 'pound', or measurements like 'ton' and 'pint'. All of the nouns that had more than ten tokens using a numerative used that numerative for at least 60% of the tokens I found. This indicates that numeratives are used fairly consistently, but it varies based on the noun and some nouns did have a low percentage of numerative use over total tokens.

Based on my data, the use of duals is decreasing over time, while the use of the numeratives identical to singular are staying consistent in use in general and increasing for

latha/là 'day' and *sgillinn* 'penny, pence'. The significant increasing pattern, as mentioned in §5.2 could reflect an actual increase in numerative use or it could be due to representation in number of texts or genre of the texts.

Unfortunately, the Canadian dialect did not have enough total tokens to say whether numeratives still are in use in that dialect and if so, how their use is changing over time. The Peripheral dialect also had a small fraction of the tokens that the Central dialect had, but in my data it appeared that dialect does not influence the use of numeratives in a significant way.

From the point of view of complexity, the consistent use of numeratives that are identical in form to a singular makes sense, even with the increase in L2 learners. The numerative form is not a unique form, so it is not an additional form that speakers must remember; instead they are just using the form in a non-typical way. For the decrease in the use of dual, it also makes sense because having a separate dual form increases complexity and having a large amount of L2 learners causes a loss of complexity in languages (McWhorter 2008: 168, Karlsson et al. 2008: viii).

Of the nouns that used a numerative identical in form to singular that were found in this project, almost a third of them only occurred once. Due to the low number of results for those words, it is hard to make any conclusions about how they are used in general. The authors of the texts in the corpus could use these nouns in singular in general or it could have been an overgeneralization, but without more examples, it is hard to know.

Due to the time restrictions of a master's thesis, it is impossible to go as in-depth with this research as could be possible. It would be beneficial to try to capture all nouns that come after numerals three through ten to look for nouns that use a numerative, as opposed to just three and four. It could also be beneficial to widen the scope of the project to be able to include more texts from the corpus by not limiting the results to texts with geographic origins and date of the language, such as using the publishing date instead of the date of the language. In the future, when there is a large corpus of spoken language available for Scottish Gaelic, it would be interesting to compare what numeratives, if any, are found.

References

Acquaviva, Paolo. 2006. Goidelic Inherent Plurals and the Morphosemantics of Number. *Lingua* 116. 1860–1887. doi: 10.1016/j.lingua.2004.10.003

Bauer, Micheal. & Robertson, William. 2016. *Am Faclair Beag.* (https://www.faclair.com/index.aspx) (Accessed 2020-04-25.)

BBC News. 2013. *Census Shows Decline in Gaelic Speakers 'Slowed'.* (https://www.bbc.com/news/uk-scotland-highlands-islands-24281487) (Accessed 2020-01-28.)

Broderick, George. 2009. Manx. In Ball, Martin & Müller, Nicole (eds.), *The Celtic Languages,* 230–304. London: Routledge.

Bybee, Joan. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10 (5). 425-455. doi: 10.1080/01690969508407111.

Christian Brothers. 1960. *Graiméar Gaeilge na mBráithre Críostaí.* Dublin: An Gúm.

Corbett, Greville. 2000. *Number.* Cambridge, UK; New York: Cambridge University Press.

Corbett, Greville, Hippisley, Andrew, Brown, Dunstan & Marriott, Paul. 2001. Frequency, Regularity and the Paradigm: A Perspective from Russian on a Complex Relation. In Bybee, Joan & Hopper, Paul (eds.), *Frequency and the Emergence of Linguistic Structure*, 201–226. Amsterdam: John Benjamins. doi: 10.1075/tsl.45.11cor

Corpas na Gàidhlig, Digital Archive of Scottish Gaelic (DASG). University of Glasgow. (https://dasg.ac.uk/corpus/) (Accessed 2020-04-24.)

Dahl, Östen. 2004. *The Growth and Maintenance of Linguistic Complexity*. Amsterdam; Philadelphia: John Benjamins.

de Bhaldraithe, Tomás. 1977. *Gaeilge Chois Fairrge: An Deilbhíocht*. Dublin: Dublin Institute for Advanced Studies.

Draskau, Jennifer Kewley. 2012. *Practical Manx.* Liverpool: Liverpool University Press.

Fenech, Edward. 1996. Functions of the Dual Suffix in Maltese. *Rivista di Linguistica* 8 (1). 89-99.

Gillies, Williams. 2009. Scottish Gaelic. In Ball, Martin & Müller, Nicole (eds.), *The Celtic Languages,* 230–304. London: Routledge.

Greene, David. 1992. Celtic. In Jadranka Gvozdanović (ed.), *Indo-European Numerals,* 497–554. Berlin; New York: de Gruyter.

Haugen, Einar. 1950. The Analysis of Linguistic Borrowing. *Language* 26. 210–231.

Jackson, Kenneth H. 1968. The Breaking of Original Long ē in Scottish Gaelic. In Carney, James & Greene, David (eds.), *Celtic Studies: Essays in Memory of Angus Matheson, 1912-1962,* 65-71. London: Routledge and Kegan Paul.

Joseph, Brian D. 1998. Diachronic Morphology In Spencer, Andrew & Zwicky Arnold M. (eds.), *The Handbook of Morphology*. Oxford: Blackwell Publishing

Karlsson, Fred & Miestamo, Matti & Sinnemäki, Kaius. 2008. Introduction: The Problem of Language Complexity. In Miestamo, Matti & Sinnemäki, Kaius & Karlsson, Fred (eds.), *Language Complexity: Typology, Contact, Change*, vii–xiv. Amsterdam: John Benjamins.

Lamb, William. 2002. *Scottish Gaelic.* 2nd edn. München: LINCOM Europa.

MacAulay, Donald. 1992. The Scottish Gaelic Language. In MacAulay, D. (ed.), *The Celtic Languages*, 137-248. Cambridge: Cambridge University Press.

MacKinnon. 1911. The Dual Number in Gaelic. *The Celtic Review,* 7(25). 1-10. doi:10.2307/30070373

McEnery, Tony, & Hardie, Andrew. 2011. *Corpus Linguistics: Method, Theory and Practice.* Cambridge: Cambridge University Press. doi:10.1017/CBO9780511981395

McWhorter, John. 2008. Why Does a Language Undress?. In Miestamo, Matti & Sinnemäki, Kaius & Karlsson, Fred (eds.), *Language Complexity: Typology, Contact, Change*, 167–190. Amsterdam: John Benjamins.

Miestamo, Matti. 2008. Grammatical Complexity in a Cross-Linguistic Perspective. In Miestamo, Matti & Sinnemäki, Kaius & Karlsson, Fred (eds.), *Language Complexity: Typology, Contact, Change*, 23–41. Amsterdam: John Benjamins.

Moseley, Christopher (ed.). 2010. *Atlas of the World's Languages in Danger*. 3rd edn. Paris: UNESCO Publishing. Online version (http://www.unesco.org/culture/en/endangeredlanguages/atlas) (Accessed 2020-01-28).

Mufwene, Salikoko S., Coupé, Christophe & Pellegrino, François. 2017. Complexity in Language: A Multifaceted Phenomenon. In Mufwene, Salikoko S., Coupé, Christophe & Pellegrino, François (eds.), *Complexity in Language: Developmental and Evolutionary Perspectives,* 1-29. Cambridge: Cambridge University Press. doi:10.1017/9781107294264

National Archives of Scotland. 2020. *2011 Census of Scotland.* Retrieved from: (https://www.scotlandscensus.gov.uk/ods-web/area.html) (Accessed 2020-04-21.)

Nilsen, Kenneth E. 2010. A' Ghàidhlig an Canada: Scottish Gaelic in Canada. In Watson, Moray & Macleod, Michelle (eds.), *The Edinburgh Companion to the Gaelic Language,* 90-107. Edinburgh: Edinburgh University Press.

Nurmio, Silva. 2019. Grammatical Number in Welsh: Diachrony and Typology. Malden, MA, Oxford and Chichester: Wiley & Blackwell.

Nurmio, Silva. & Willis, David. 2016. The Rise and Fall of a Minor Category: The Case of the Welsh Numerative. *Journal of Historical Linguistics* 6(2). 297–339.

Ó Baoill, Dónall P. 1999. *An Teanga Beo: Gaeilge Uladh*. Dublin: Institiúid Teangeolaíochta Éireann.

Ó hAnluain, Liam. (ed.) 1999. *Graiméar Gaeilge na mBráithre Críostaí*. Dublin: An Gúm.

Ó Maolalaigh, Roibeard. 2013. Corpas na Gàidhlig and Singular Nouns with the Numerals 'Three' to 'Ten' in Scottish Gaelic. In Anderson, Wendy (ed.) *Language in Scotland: Corpus-Based Studies*, 113-142. Amsterdam: Rodopi.

Ó Sé, Diarmuid. 2000. *Gaeilge Corca Dhuibhne*. Dublin: Institiúid Teangeolaíochta Éireann.

Ó Siadhail, Micheal. 1982. Cardinal numbers in Modern Irish. *Ériu* 23. 101-107.

Ó Siadhail, Micheal. 1989. *Modern Irish.* Cambridge: Cambridge University Press.

Ross, Susan. 2016. *The Standardisation of Scottish Gaelic Orthography 1750-2007: A Corpus Approach.* Glasgow: University of Glasgow. (PhD thesis). (http://theses.gla.ac.uk/7403/) (Accessed 2020-03-06.)

Wagner, Heinrich, 1964. *Linguistic atlas and survey of Irish dialects. Vol. II. The dialects of Munster.* Dublin: Dublin Institute for Advanced Studies.

Wagner, Heinrich, 1966. *Linguistic Atlas and Survey of Irish Dialects. Vol. III. The Dialects of Connaught*. Dublin: Dublin Institute for Advanced Studies.

Wagner, Heinrich & Ó Baoill, Colm. 1969. *Linguistic Atlas and Survey of Irish Dialects. Vol. IV. The Dialects of Ulster and the Isle of Man. Specimens of Scottish Gaelic Dialects. Phonetic Texts of East Ulster Irish.* Dublin: Dublin Institute for Advanced Studies.

Appendix

Due to the large amount of results, it is impossible to attach all of my tokens here. To view the spreadsheets of results, use this link: https://zenodo.org/record/3817441#.XrV5a2gzaUk. Table 20 is a small example of what the collected corpus data looks like.

Table 20: Example of collected corpus data

Search Result	Source	Date of langu	Geographic orig	Noun	Singular?	Dialect
na gréine ann an ceithir bliadhna	113	1850-1899	Argyll	bliadhna	Y	Central
ann an ceithir bliadhna	113	1850-1899	Argyll	bliadhna	Y	Central
an tri bliadhna — cuid ann an sea	113	1850-1899	Argyll	bliadhna	Y	Central
tha ochd - làithean - fichead re thri bliadhna	113	1850-1899	Argyll	bliadhna	Y	Central
Mhair e mar so , rè ùine thrì bliadhna	118	1800-1849	Caithness	bliadhna	Y	Peripheral
a shìde car thrì bliadhna aig Comunn na Stuar	124	1800-1849	Perthshire	bliadhna	Y	Peripheral
a thug corr is tri bliadhn ' eug thar fhichead	129	1800-1849	Lochaber	bliadhna	Y	Central
n chosg e tri bhiadhn ' eug thar fhichead	129	1800-1849	Lochaber	bliadhna	Y	Central
bhliadhna gu ceithir bliadana dh ' aois	136	1800-1849	Argyll	bliadhna	Y	Central
dh ' fhan e maille riu car tri bliadhna	145	1800-1849	Perthshire	bliadhna	Y	Peripheral
no eadhon 3 bliadhna	196	1800-1849	Gairloch	bliadhna	Y	Central
ni ' s lugha na 3 bliadhna	196	1800-1849	Gairloch	bliadhna	Y	Central
an sin airson trì bliadhna a ' coimhead às dea,	197	1950-1999	Easter Ross	bliadhna	Y	Peripheral
Agus cha robh e trì bliadhna pòsta dar chaoch	197	1950-1999	Easter Ross	bliadhna	Y	Peripheral
soluis ceithir bliadhna air an t - slighe	201	1900-1949	Lewis	bliadhna	Y	Central
Sgoil Steòrnabhaigh airson trì bliadhna eile	201	1900-1949	Lewis	bliadhna	Y	Central
an déidh trì bliadhna gu leth ri aghaidh nàmh,	201	1900-1949	Lewis	bliadhna	Y	Central
Trì bliadhna a bha e air falbh	202	2000-	Lewis	bliadhna	Y	Central

Table 21 Results per category for d[h,][à,a,á] and '2'

Category	d[h,][à,a,á]	'2'
Or Phrase	118	13
Complex Number	933	0
Standard	3522	117

Of the Or Phrase tokens found in these searches, five tokens do not follow the typical pattern, three tokens from the written form search and two tokens from the numeric form search. For the Complex Number category from the *dà* 'two' search, five tokens do not follow the typical pattern for Scottish Gaelic. Out of 3522 tokens in the Standard category for the written form, 185 are dual tokens, consisting of 47 unique nouns (seen in Table 5), and the rest are singular, which is the typical pattern in Scottish Gaelic. For the numeric form, one token from the Standard category uses a dual form. See §2.4 for a discussion on number in Scottish Gaelic.

Table 22: Results per category for t[h,]r[ì,i,í], '3', c[h,][e,è,é][i,ì,í]th[i,ì,í]r, and '4'

Category	t[h,]r[ì,i,í]	'3'	c[h,][e,è,é][i,ì,í]th[i,ì,í]r	'4'
Or Phrase	182	29	44	16
Complex Number	514	2	435	0
Standard	1249	79	487	51

For the Or Phrase category, 21 tokens do not follow the typical pattern. Of the complex numbers, 67 of the total tokens do not follow the typical pattern for Scottish Gaelic. For the Standard Category, out of 1866 total tokens, 490 are numerative tokens. These 490 numerative tokens consist of 105 unique nouns. The rest of the tokens are singular, which is the typical pattern in Scottish Gaelic. See §2.4 for a discussion on number in Scottish Gaelic.

Results for Or Phrases

There were a total of 196 Or Phrase tokens from all of my searches. Of these, 21 tokens were Or Phrases that did not follow the typical pattern, where numerals three through ten are followed by a plural noun.

Table 23: Or Phrases that have Numeratives

Or Phrases	Source	Date of Lang	Geo. Origin	Noun	Dialect
dhà no trì cheudan bliadhna	201	1900-1949	Lewis	*ceud*	Central
dhà no trì cheudan mìle	201	1900-1949	Lewis	*ceud*	Central
2 no 3 bliadhna	196	1800-1849	Gairloch	*bliadhna*	Central
2 no 3 bliadhna	196	1800-1849	Gairloch	*bliadhna*	Central
trì no ceithir bhliadhna	197	1950-1999	Easter Ross	*bliadhna*	Peripheral
thrì no cheithir bliadhna	106	1850-1899	Lewis	*bliadhna*	Central
ceithir no còig bliadhna	500005	2000-2016	Uist	*bliadhna*	Central
3 no 4 chuairt	196	1800-1849	Gairloch	*cuairt*	Central
3 no 4 chuairt	196	1800-1849	Gairloch	*cuairt*	Central
3 no 8 tastan	196	1800-1849	Gairloch	*tastan*	Central
3 no 8 tastan	196	1800-1849	Gairloch	*tastan*	Central
dhà no thrì thastan	197	1950-1999	Easter Ross	*tastan*	Peripheral
tri no ceathair latha	38	1950-1999	Harris	*latha*	Central
tri no ceathair latha fhad	38	1950-1999	Harris	*latha*	Central
da no tri chrodh	196	1800-1849	Gairloch	*crodh*	Central
trì no ceithir mìle	197	1950-1999	Easter Ross	*mile*	Peripheral
Dìreach dhà neo trì diofar	202	2000-2016	Lewis	*diofar*	Central
trì no ceithir uair	197	1950-1999	Easter Ross	*uair*	Peripheral
trì no ceithir uair	197	1950-1999	Easter Ross	*uair*	Peripheral

| *cheithir no chóig urad* | 201 | 1900-1949 | Lewis | *urad* | Central |
| *cheithir no chóig urad* | 201 | 1900-1949 | Lewis | *urad* | Central |

The two tokens where *ceudan* 'hundreds' (singular *ceud* 'hundred') are in this table because the typical pattern is that *ceud* 'hundred' would be in singular following *dà* 'two' or *trì* 'three'. But as seen above, in these two Or Phrases, there is a plural form instead.

For *bliadhna* 'year', five of the 19 tokens use a numerative identical in form to a singular. These five tokens come from four different sources and four different time periods. *Bliadhna* 'year' is used with a numerative form in high frequency in my data with three and four (97.02%, Table 10) so it is unsurprising that it also appears using a numerative in or phrases as well. *Tastan* 'shilling' uses a numerative identical in form to a singular in three of six Or Phrases. The three tokens come from two different sources and two different time periods. *Latha* 'day' uses a numerative identical in form to a singular for two out of twenty Or Phrases, but both of the numerative tokens are from the same source. *Cuairt* 'time, cycle' uses a numerative identical in form to a singular twice out of two Or Phrases. However, both tokens are from the same source. *Uair* 'hour' also had two tokens that use a numerative identical in form to a singular with both coming from the same source. *Uair* 'hour' appeared in 17 Or Phrases. Finally, *urad* 'number, quantity' had two Or Phrases, which were both from the same source and use a numerative identical in form to a singular.

Results for Complex Numbers

Out of all the tokens in my data that contained Complex Numbers, 12 out of 1884 used a non-typical form on either the second number noun (such as *fichead* 'twenty' or *ceud* 'hundred') or on the non-number nouns (meaning *bliadhna* 'year' in *trì fichead bliadhna* 'sixty years').

Complex Numbers	Source	Date of Lang	Geo. Origin	Noun	Dialect
dà shlait diag	74002	1900-1949	Western Isles	*slat*	Central
dà shlait dhiag	74002	1900-1949	Western Isles	*slat*	Central
dà làimh dhiag	49	1950-1999	South Uist	*làmh*	Central
dà luing dheug	137	1800-1849	Lochaber	*long*	Central
dà luing dheug	106	1850-1899	Lewis	*long*	Central
trì ' cheudan troidh	108	1800-1849	Argyll	*ceud*	Central
trì muilleinean	108	1800-1849	Argyll	*muillean*	Central
trì mhìltean	106	1850-1899	Lewis	*mìle*	Central
trì muilleanan	106	1850-1899	Lewis	*muillean*	Central

trì Ficheadan	202	2000-	Lewis	*fichead*	Central
trì ficheadan	500004	2000-	Uist	*fichead*	Central
ceithir muillionan	62	1900-1949	Harris	*muillean*	Central

Of these non-typical Complex Number phrases, two are from 1800-1849, two are from 1850-1899, three are from 1900-1949, one is from 1950-1999, and two are from 2000-2016. In three of the ten non-typical results, the non-number noun has a plural ending and in seven of the ten non-typical results, the second number word has a plural ending. All of the non-typical tokens from the *dà* 'two' searches, use a dual form with 'twelve', where a singular is expected, but two of these are from the same text. Most of the non-typical Complex Number tokens are used with 'three', which is six of the ten tokens. All of the non-typical Complex number tokens are from the Central dialect.