Linguistic densification in web-based English: comparing grammatical features of density across registers

M van Zyl

orcid.org/0000-0002-6737-5929

Dissertation accepted in fulfilment of the requirements for the degree Master of Arts in Linguistics and Literary Theory at the North-West University

Supervisor: Dr YV Botha

Co-supervisor: Prof AJ van Rooy

Graduation: August 2021

Student number: 23629975

ACKNOWLEDGEMENTS

All the honour and glory to the Lord for giving me the ability and support structure to have completed this study.

I would also like to thank the following people for their support and contributions:

My supervisor, Dr Lande Botha, without whom I would not be a linguist. Her patient guidance, intellect, understanding and encouragement are unqualified;

My co-supervisor, Prof Bertus van Rooy, for his support, forbearing character, and also for the vigour and flair in his approach;

My husband, Sternberg, for patiently carrying the burden of our responsibilities due to my studies;

My family and friends for their support, but in particular, Wilma Davel, for her prayers over the last twenty years;

The Research Unit, CTexT and the Ferdinand Postma Library, in particular Mr Wikus Pienaar, Ms Bernice Mackenzie, Ms Elsa van Tonder, Ms Nicoline Gerber, Ms Anneke Coetzee and Ms Helah van der Waldt, for their respective personal and academic support; and

The North-West University, the Research Unit as well as the Department of Sport, Arts and Culture, whose contributions have provided much appreciated financial relief.

ABSTRACT

This dissertation is an investigation of grammatical patterns related to the density of noun phrases in online and offline registers of English. Noun phrase density involves embedding phrasal and clausal modifiers and complements into the structure of the noun phrase, creating informationally dense and complex grammatical structures. Research shows that written registers, and especially registers with information-driven purposes, have higher relative frequencies of density devices, where popular nonfiction and fiction have lower relative frequencies. Features of noun phrase density are attested in offline registers. Whether similar patterns hold for web-based or online registers is unknown. Where the grammatical realisations of discourses are influenced by situational factors, the advancement of technology and the need to produce more information also affect the lexicogrammatical choices users make when constructing language in either the spoken or written form. People have recourse to the same grammatical system in an online, that is, web-based, context as they do in an offline context. Given the online mode, this study therefore investigates the extent to which the choice to create grammatically dense noun phrases is exercised in an online environment compared to offline situations. Five offline (COCA) and five online (CORE) registers were selected to help answer this question. These registers vary in their communicative purpose. A nonwritten spoken component is also included in the analysis. The analysis emphasises three levels of contrasts: spoken versus written language, online versus the offline written form, and pairwise comparisons between registers. The pairwise comparisons are based on a percentage difference (effect size) metric to quantitatively describe the differences between these varying registers. The results suggest that, as a whole, online registers have proportionally more densification than offline registers do. Finer grained differences in frequencies emerge that largely depend on the type of density device. The results of the percentage difference metrics suggest that the spoken component is not that different from some written registers, and that even though online registers have higher proportions, these registers are also similar to offline registers in various respects relating to density.

Key terms: grammatical density, densification, noun phrase modifier, noun phrase complement, web-based English, online register, offline register

OPSOMMING

Hierdie verhandeling is ’n ondersoek na die gebruik van grammatikale patrone wat verband hou met die digtheid van naamwoordstukke in aanlyn en vanlyn Engelse registers. Naamwoordstukverdigting behels die inbedding van bepalers en komplemente in die naamwoordstuk en gee aanleiding tot inligtingsdigte en grammatikaalkomplekse strukture. Navorsing toon aan dat geskrewe registers en meer spesifiek registers met 'n inligtinggedrewe doel, hoër relatiewe frekwensies van digtheidpatrone het, teenoor meer algemene en populêr-geskrewe registers, soos fiksie en niefiksie, wat laer relatiewe frekwensies in die gebruik van hierdie patrone toon. Daar is reeds bewyse vir naamwoordstukverdigting in vanlyn registers. Of die gebruik van hierdie patrone in aanlyn registers eens is met vanlyn registers, is nie bekend nie. Terwyl konteks ’n rol speel in die linguistiese realisering van diskoers, dra tegnologiese ontwikkelinge, asook die verandering in inligtingsbehoefte, by tot die leksiko-grammatikale taalkeuses wat uitgeoefen word. Mense het juis toegang tot dieselfde linguistiese sisteem ongeag die diskoerskonteks, hetsy in die aanlyn of vanlyn konteks. Gegewe die aard van die aanlyn medium probeer hierdie studie die mate waartoe digtheidsstrukture gebruik word in die aanlyn konteks teenoor ’n vanlyn konteks ondersoek. Vyf vanlyn en vyf aanlyn registers is gekies om analises uit te voer. Hierdie registers wissel in terme van hul kommunikatiewe fokus en doel. ’n Niegeskrewe gesproke komponent word ook ingesluit. Die analise beklemtoon drie vlakke van vergelyking: gesproke teenoor geskrewe taal, aanlyn teenoor die vanlyn geskrewe vorm, en vergelykings tussen pare registers. Die vergelykings tussen pare registers word uitgevoer deur te steun op ’n persentasie verskil (effekgrootte) toets, waar verskille tussen verskeie registers kwantitatief beskryf word. Dit blyk uit die resultate dat aanlyn registers oor die algemeen ’n hoër gebruik van verdigting toon as vanlyn registers. Meer spesifieke verskille in relatiewe frekwensies tree egter na vore afhangende van die tipe digtheidpatrone. Verder dui effekgrootte aan dat die gebruik van hierdie strukture nie baie verskillend is tussen die gesproke komponent en sommige geskrewe registers nie, en dat, alhoewel die tendens aantoon dat hierdie patrone proporsioneel meer voorkom in aanlyn registers, aanlyn en vanlyn registers in sommige opsigte dieselfde in terme van digtheid is.

Sleutelterme: grammatikale digtheid, verdigting, naamwoordstukbepaler, naamwoordstukkomplement, web-gebaseerde Engels, aanlyn register, vanlyn register

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 CONTEXTUALISATION AND PROBLEM STATEMENT

1.1 Introduction

The pressures to communicate information in a concise and efficient way, together with factors such as that of a specialised readership and text purpose, are functional motivations for a type of grammatical complexity known as densification (Leech, Hundt, Mair & Smith, 2009:210). This chapter introduces the aspects that relate to density through nominal modification and complementation in web- and nonweb-based registers. Various studies focus on density and noun phrase modification and complementation patterns over time and in different registers (Biber & Clark, 2002; Biber & Gray, 2011; Gray, 2015; Leech *et al*. 2009). With the introduction of the information age however, the current status of noun phrase density as information compression device is unknown. Further, the web as language source have been employed in a variety of linguistic investigations, but only a select few focus on the categorisation and characterisation of online registers and register differences in English (Biber & Egbert, 2016; Egbert, Biber & Davies, 2015; Johnson, 2015). Biber and Clark (2002), Biber and Gray (2012) and Crystal (2011) argue that the internet results in the emergence of complex linguistic situations and discourses. Further, multidimensional analyses (MD analyses) have shown that nouns and some elements surrounding noun phrase modification generally distinguish spoken language from the written form, but also information-driven texts from more general writing. This study aims to investigate a specific subset of modification and complementation features that create dense noun phrases in a range of information-driven and general nonfiction and fictional registers. Current distribution patterns of nouns and noun phrase elements suggest that written discourse, and especially those with an information purpose, have higher frequencies of many of these features, compared to general writing and conversation (Biber, Johansson, Leech, Conrad, & Finegan, 1999). This chapter then provides a framework for the investigation of density in web-based and nonweb-based registers, by broadly describing the types of structural embedding involved in density and briefly outlining the role the web has played in linguistic inquiry. Finally, the research question, aims and methodology are formulated in light of this contextualisation.

1.2 Contextualisation and problem statement

According to Biber (2001:215) discourse types or registers have been thought to be grammatically complex because of the presence of dependent clauses. However, linguistic multidimensional analysis (MD analysis) show that nouns and some features surrounding noun phrase modification often co-occur in those registers that are assumed to be grammatically complex, and not so much dependent clauses (Biber, 1988). Based on findings from MD analysis (which distinguish registers

based on underlying communicative purposes), Biber and Gray (2011, 2016) address the assumption that complexity involves structural elaboration (through dependent clauses) by distinguishing between different types of grammatical complexity. These are elaborated structure and condensed structure.

The first is a style that relies extensively on the use of dependent clauses or clausal modification; that is, modification through elaborative structures on clause level, while the condensed structure employs phrasal and particularly noun phrasal modification, that is, embedded phrases (Biber & Gray, 2011:226), but also embedded clauses (Gray, 2015). Nouns can be modified by both phrases and clauses and Gray (2015:65) argues that clausal forms such as relative clauses are embedded at phrasal level despite serving an elaborative function. This implies that modification in the noun phrase is primarily an information compression strategy, which aligns with the way in which Leech *et al*. (2009) employ the term densification.

The following text examples illustrate the distinction between the discourse styles Biber and Gray (2016:16-17) want to emphasise. The first example (1) shows a main clause with multiple dependent clauses. The second example (2) has fewer main clause elements, or a simple clause structure (one verb phrase), but a complex noun phrase, with multiple embedded phrasal postmodifiers. The elements of the main clause in both example (1) and (2) are bracketed in pink. Postmodifiers are underlined in the examples below, where the common nouns that are postmodified are in bold. Premodifiers are in italics. In (1), the postmodifier is in the form of a clause and in (2) in the form of a phrase. In example (1), the elements of the various dependent clauses are in different colours, namely green, purple, black and red:

1. [It] [was] [just] [one of those **things** [where [I] [think] [[Paul] [’s gotten] [to [the **point** [where [he] [won’t just go on and accept] [[what [she] [says]]]]]]]. (Biber & Gray, 2016:16).

2. [The *distinctive* **effect** [of [the **size** [of [the *Asian* **population** [on [*income* inequality]]]]]] [certainly] [deserves] [further research]. (Gray, 2015:50).

These contrasting examples of complex texts are the reason why Biber and Gray (2016) challenge the traditional view of informational writing as structurally elaborated and therefore grammatically complex. The first is an example illustrating this structural elaboration; that is, the addition of elements on a clausal level. The modification devices used in the second example have the function of compressing information within a single clause element (the subject in this case). Biber and Gray (2016:18) emphasise that:

> these devices [from the second example] constitute a second major type of grammatical complexity,

> not associated with structural elaboration. In fact, they have exactly the opposite function: maximally compressing structure rather than elaborating structure.

Compressed discourse complexity is further characterised by Gray (2015:54) as the following:

> the information [that] is added to noun phrases in optional phrases that can be considered more condensed alternatives to fuller clausal structures. Features like prepositional phrases and nouns as nominal pre-modifiers convey meanings that could be more explicitly stated through elaborating clausal structures.

This study deals with this second major type of grammatical complexity and its associated features with regard to noun phrase compression. Embedding elements in the noun phrase is considered to be synonymous with information compression strategies. This strategy is what Biber and Gray (2012:326) refer to as economy. Leech et al. (2009:210) refer to this as densification. Density ultimately involves the use of embedded structures within the noun phrase, where the alternative would be to use multiple clausal structures to express the same information.

Density is traced from the 18th century up until the late 20th century (Biber & Gray, 2011; Leech *et al*., 2009). Nouns and noun phrase structures do not mark differences between more general or popular writing and informational writing in 18th century English (Biber & Gray, 2011). This changes however from the 19th century into the 20th century, where popular prose and specialist writing start to diverge in terms of the frequency of these structures (Biber & Gray, 2011). Interestingly however, Leech *et al*. (2009) show that information-driven registers as well as popular fiction and nonfictional registers all have increased frequencies of noun phrase structures by the late 20th century. According to Biber and Clark (2002:44) an "information explosion" is largely responsible for these changes over time. The general findings from these studies provoke an investigation into the frequencies of noun phrase modifying and complementizing structures in contemporary web-based English in comparison with nonweb-based registers.

Advocates of the web as corpus promote the internet as a source for language inquiry, which, in its size and recentness, enables researchers to adequately study grammatical phenomena that would otherwise (in traditional corpora) be impossible (Rosenbach, 2007:182). The web has motivated multidisciplinary research within both communicative and linguistic domains and has been an important source of information for studies on language variation and change (Anderwald, 2007; Baron, 2008; Beißwenger, 2008; Davies, 2013; Herring, 2013; Herring, 2010; Lüdeling, Evert & Baroni, 2007; Mair, 2007; Rohdenburg, 2007; Seargeant & Tagg, 2011; Squires, 2010; Tagliamonte, 2016; Tannen, 2013; West & Trester, 2013).

The web as language source has also probed research into register variation. Biber and Egbert (2016) and Egbert, Biber and Davies (2015) performed a categorisation of texts found on the web.

Biber and Egberg (2016) followed this by a factorial analysis of these registers (of which a subset forms part of this study in particular). 17% of texts were categorised as having an informational purpose, where 22.5% of these were categorised under "description of a thing" (Biber & Egbert, 2016:104). Narrative texts made up 31% of texts and spoken only 0.9% (Biber & Egbert, 2016:101). Additionally, a multitude of hybrid-register classifications surfaced. Their MD analysis finally yielded nine dimensions, showing that web-based registers confirm and reflect language variation and that these dimensions are quite diverse in their communicative purpose compared to dimensions some previous MD studies have shown (Biber & Egbert, 2016).

Further, Crystal (2011:17) specifically argues that the online mode influences the selection of grammatical choices exercised in linguistic contexts. Language users on the web have recourse to the same system of lexicogrammatical choices as they would have in offline communication. Biber and Clark (2002) and Biber and Gray (2012) add that these advancements encourage the demand to produce information. The natural inference that technology (in addition to contextual factors) will impact grammatical choices to some extent can be drawn. The advancement of technology paired with an increase in information needs, may induce grammatical choices relating to compressed language usage across web-based registers. The aim of the current study is therefore to investigate the linguistic realisation of density in web-based registers of English compared to offline registers. The structural embedding strategies involved in information density, may play out differently for registers on the web, given the mode – and therefore the nature – of the production and discourse constraints.

Structural embedding, as an information compression strategy, has been the focus of a variety of investigations of corpora, including the Brown family of Written Standard English (Leech & Smith, 2009; Leech *et al.*, 2009), the Representative Corpus of Historical English Registers corpus (ARCHER) (Biber & Clark, 2002) and the Longman Corpus of Spoken and Written English (Biber *et al.*, 1999). Densification strategies and their changing patterns are diachronically accounted for in traditional English corpora (Biber, 2003; Biber & Clark, 2001; Biber & Gray, 2011). However, the introduction of the information age has brought with it complex modes of additional communication channels, and which arguably make the distinction between spoken and written discourse less distinct (Halliday, 1989:81). In addition to this, Crystal (2006:5) emphasises that the online medium effects the communicative context in its entirety. He states that the electronic medium is fundamentally different from other semiotic situations, and this ultimately both constrains and facilitates the online context in a way that is different from conventional linguistic communicative circumstances (Crystal, 2006:5). It is not clear firstly, to what extent density is a distinguishing feature of web-based registers of English (in comparison with nonweb-based registers), and secondly, if it sharply distinguishes contemporary spoken and written forms as it does with historical data. While densification is a prominent feature of written language, Biber and

Gray (2012) emphasise that all register types, including spoken language, employ densification strategies to some extent. In part, this study attempts to respond to this by analysing density features in a variety of mostly written registers but with the inclusion of a spoken component.

Exercising the choice of one grammatical structure over another (embedded structures as opposed to using multiple higher clause level elements) is influenced by a complex combination of contextual, production and processing constraints. Additionally, discourse motivational factors largely determine the preference of certain grammatical structures over others, which influence the grammatical characteristics of various written and spoken registers (Biber & Conrad, 2009; Biber & Conrad, 2003; Biber & Conrad, 2001; Cutting, 2013). These include, among various considerations, the purpose for which a text is constructed, the audience for whom the text is intended and the real-time pressure under which language is produced (Biber & Conrad, 2009); not all text-types are subjected to the same constraints and this will influence the final linguistic product.

It was previously mentioned that most written registers employ features of density, but that the degree to which these features occur in registers largely depends on the situational context on which a text is built. Mode as an additional functional motivation to general communicative purposes may also influence the grammatical choices employed in linguistic constructions on the web. With the noun phrase and elements surrounding noun phrase modification and complementation as main distinguishing linguistic feature of density, this study focuses on how different web-based registers are characterised by the need to present experiences as products, given their web-based origins. In other words, to which extent are web-based registers participating in noun phrase density, and how are they characterised by these features? These questions are refined later in this chapter.

Further, research in register variation shows that nouns and elements surrounding noun phrase modification are largely responsible for distinguishing spoken registers from the written form, but also involved or popular writing from those registers with an information-driven purpose (Biber, Johansson, Leech, Conrad & Finegan, 1999; Biber, 1988; Halliday, 1989). Density features are shown to be much more frequent in registers with an informational purpose as opposed to popular and involved text types (Biber & Gray, 2012; Biber & Gray, 2011; Gray 2015). Biber (2003), for instance, compare the frequencies of nominal pre- and postmodifiers across conversation, fiction, newspaper articles and academic discourse, and find that news and academic writing contain visibly higher frequencies of these features than general popular and involved registers do. These findings are often mirrored in most studies focused on the distribution patterns of noun phrase modification (Biber *et al*., 1999; Biber & Clark, 2002; Biber & Gray, 2012; Biber & Gray, 2011).

These findings also inform the general expectations with regard to relative frequency patterns of noun phrase modifiers and complements in the data here.

1.3 Research question and aims

Existing studies of web-based language usage have focused on noticeable linguistic features such as the use of abbreviations and colloquialisms, while no attempt has been made to characterise web-based English in terms of noun phrase density, by way of a statistical comparison of a coherent set of linguistic features with non-web-based language. Therefore, the overarching question addressed in this study is:

> how do various web-based and nonweb-based registers compare in their density, as expressed through categories of noun modification and complementation?

Assuming register differences, the study therefore aims to assess whether web-based language is influenced by grammatical density to the same extent nonweb-based language is.

1.4 Methodology

The analysis compares the normed frequencies of density features in terms of three main levels of comparison between the ten registers of web-based and nonweb-based English. Firstly, a pairwise comparison of each feature between all of the registers is made, which allows for an overall data set where all ten registers are compared simultaneously. Therefore, a web-nonweb distinction is made in lieu of the research question stated earlier. %DIFF (percentage difference) is used as an effect size measure to compare the normed frequencies of the various modifier and complementizer types in selected web-based and nonweb-based writing and a spoken register.

The second level of comparison focusses on web and nonweb written registers compared to Spoken. Given that a spoken component is included in the register selection, a comparison between the spoken register and other written web and nonweb registers is possible. Reference in terms of the differences (and similarities) between Spoken and other web and nonweb registers is imperative, because major discourse distinctions lie in differences between spoken and written language. The web hosts an array of spoken-like discourse in the written form and to not include this would leave the analysis wanting.

Further, not all corpora allow for cross-register comparison. The registers in the COCA and CORE corpus are specifically constructed in such a way that they allow for sampling across individual registers and by extension allow one-to-one or one-to-all comparisons, effectively measuring the differences and similarities across these sets. However, while this may lead to some frequency expectations between seemingly similar registers (only in name) such as web-news and nonweb-

news, assumptions about similarity remain so until they are verified. Since little is known about the construction of web-based registers, a subset of the CORE corpus differing along the cline of discourse purposes, that is, narrative, informational et cetera, is selected and compared overarching to nonweb-based register in COCA, with a similar discourse variety. This includes the nonweb Spoken register.

The data comparisons therefore allow for two types of comparison, but which are complemented by %DIFF as the effect size measure. Effect size is used here to discuss the differences and similarities between registers in terms of their magnitude. The %DIFF is calculated for each individual densification device in each of the ten registers. The choice to use %DIFF (and closely related ratio scores) to describe the differences in frequency is discussed in the methodology chapter.

1.5 Prospectus

The first chapter provided a literature overview that has served as an introduction to studies of web-based linguistic inquiry, register variation and noun phrase density. Through this contextualisation, the aims and research question of this study have been set out. Chapter 2 relies on reference grammars to provide the grammatical framework for the analyses by focusing on the noun phrase and the embedded structures that are responsible for creating dense and complex noun phrases. The framework for analysis sets the foundation for the way in which densification features are identified in the corpora. Chapter 3 discusses the corpus-based approach in which concordances are sampled from the corpora and analysed according to the framework set out in Chapter 2. Further, the methodology explains the descriptive statistics and effect size measures involved in answering the research question. The research question is answered in terms of three levels of comparison. These levels of comparison are central to the qualitative analysis in Chapter 4. Chapter 4 describes the results by firstly comparing the relative frequencies of each density feature across ten registers in terms of the written registers and the spoken component. It then focuses on the differences and similarities between web-based and nonweb-based registers with reference to relative frequency distributions and the effect size magnitude of those comparisons. The study concludes with Chapter 5, where main findings and some final points relating density in web-based in contrast to nonweb-based English registers are summarised. This chapter also emphasises some other areas for possible future exploration.

CHAPTER 2 GRAMMATICAL FRAMEWORK FOR ANALYSIS

2.1 Introduction

This section aims to provide an overview of the structure of noun phrases as set out in standard reference grammars, in particular the *Longman Grammar of Spoken and Written English* (Biber *et al*., 1999, henceforth *LGSWE*) and the *Cambridge Grammar of the English Language* (Huddleston & Pullum, 2002). This is not to say that these are the only sources for information on noun phrase structure and function, but they do form the basis according to which the patterns for these analyses are identified. Finally, the patterns serve the purpose of arriving at search strings for each of the selected grammatical structures involved in prenominal and postnominal modification and postnominal complementation. These search strings are based on syntactic forms from which the modifier functions are ultimately derived. For the sake of completeness, reference will also be made to those constructions that were not included for quantitative analyses.

2.2 Noun phrase structure

According to Biber *et al*. (1999:574) and Payne and Huddleston (2002:326), noun phrases contain at least a head element that can either be a common noun, proper noun or pronoun. Occasionally, the noun phrase may also be headed by a fused head, which means that the head of the phrase is not explicitly filled by a noun, but is rather realised jointly by merging the noun and a premodifier or determiner (Payne & Huddleston, 2002:332). In this case the adjective is functioning as the head of the noun phrase as in (1) and (2) below (Biber *et al*., 1999:64):

1. the rich

2. the impossible

This study is concerned with noun-headed noun phrases and not those headed by pronouns or fused heads.

Further, apart from the head element, noun phrase structure has allocated constituent spaces for determiners that specify the head referent, premodifiers (that is, modifiers that precede the head noun) and postmodifiers, which qualify the meaning of the noun (Biber *et al*., 1999:574–575). This basic abstract structure of the noun phrase may therefore be presented as in this table below, which shows the functional slots within the noun phrase. The examples are taken from Biber *et al*. (1999:574).

Table 2-1: Functional slots in the noun phrase

Determiner	(Premodifier)	Head	(Postmodifier)
the	*small wooden*	*box*	*that he owned*
			Complement
the		*fact*	*that I haven't succeeded*

The only obligatory element is the head noun constituent, while the functions of premodifier and postmodifier are optional to fill in any particular communicative context. The presence of these modifier slots is what allows language users to introduce elaborating information into the noun phrase without creating additional clause elements to do so. Therefore, the modifier slots are responsible for allowing information-dense noun phrases. This study focuses on analysing the use of the various constructions that fill these functional slots.

Payne and Huddleston (2002:329) suggest an alternative hierarchical structure that recognises an intermediate category called the nominal, which is found between the head noun and the noun phrase. Payne and Huddleston (2002:329) exemplify the distinction between noun phrase and nominal (the italicised words form the nominal):

3. the *old man*

4. that *book you were talking about*

5. the *man*

Nominals may either be filled by nouns accompanied by dependents as in (3) and (4), or single nouns as in (5) (Payne & Huddleston, 2002:329). In (3), the nominal *old man* acts as the head of the entire noun phrase *the old man*. In turn, the noun *man* heads the nominal which is also the ultimate head of the entire noun phrase. In these examples, the dependents are the modifiers *old* and *you were talking about*. Where a nominal consists of a single noun as in (3), it is both the noun and nominal, in this case *man* (Payne & Huddleston, 2002:329).

For Payne and Huddleston (2002:330–331), this structure allows one to make a distinction between constituents that are immediate to the nominal and those that are immediate to the noun phrase. Constituents immediate to the nominal are said to be internal dependents of the ultimate

head (Payne and Huddleston, 2002:330). These elements serve as modifiers and complements of the head noun.

The dependents that are not internal to the nominal but rather immediate to the noun phrase are termed external dependents and these include elements such as quantifiers and determinatives found in the determiner function (Payne & Huddleston, 2002:330). According to Biber *et al*., (1999:588) determiners, genitives and numerals all serve a specifying rather than modifying function in noun phrases. In other words, they do not provide semantic information regarding the general qualities of the head in a noun phrase.

The distinction between external and internal dependents is useful here, since it helps narrow the focus of this analysis to the investigation of immediate pre- and postnoun dependents of the nominal. External dependents will be excluded, since they do not immediately modify the ultimate head of the noun phrase.

Payne and Huddleston (2002:439) further differentiate between internal dependents in terms of their functions. These functions are that of modifier and complement. Biber, Conrad & Leech (2002:458) define modifier as 'an omissible form that specifies further meaning about the head of the phrase', and a complement as 'a phrase or a clause that completes the meaning required by some other form' (Biber, Conrad & Leech, 2002:456).

While Biber *et al*. (1999:645) apply the distinction between complements and modifiers only to postnoun *that*- and *to*-infinitive clauses, Payne and Huddleston (2002:439) stipulate that this difference applies to both pre- and post-head constituents. Many of the complement criteria Payne and Huddleston (2002:441–443) put forth are based on semantic considerations. According to Payne and Huddleston (2002:440), licensing is the strongest argument for regarding some post-head preposition phrases as complements. This means that the selection of a preposition phrase as post-head dependent will depend on the options the noun phrase allows, as in (7) below. Payne and Huddleston (2002:441) assert that if the noun denotes a property, relation, process or action and the dependent is an involved entity, then the dependent is a complement and not a modifier (Payne & Huddleston, 2002:441).

The following sets of examples from Payne and Huddleston (2002:439) illustrate the distinction between the function of the modifier and that of the complement for both pre- and postnoun positions.

6. a <u>first-year</u> **student** [modifier]

7. a <u>linguistics</u> **student** [complement]

8. a **report** in the paper [modifier]

9. a **report** on the crash [complement]

The following pairs of examples illustrate the discrepancies in distinguishing between complements and modifiers. According to Biber *et al*. (1999:636–637), the underlined constituents in examples (8) and (9) are both postmodifiers. These are readily comparable to Pullum and Huddleston's (2002b:660–661) examples of the complement function in (10) and (11):

10. his most wounding **attack** on the tabloids

11. one apparently attractive **answer** to that question

12. an **attack** on my honour

13. an **answer** to the question

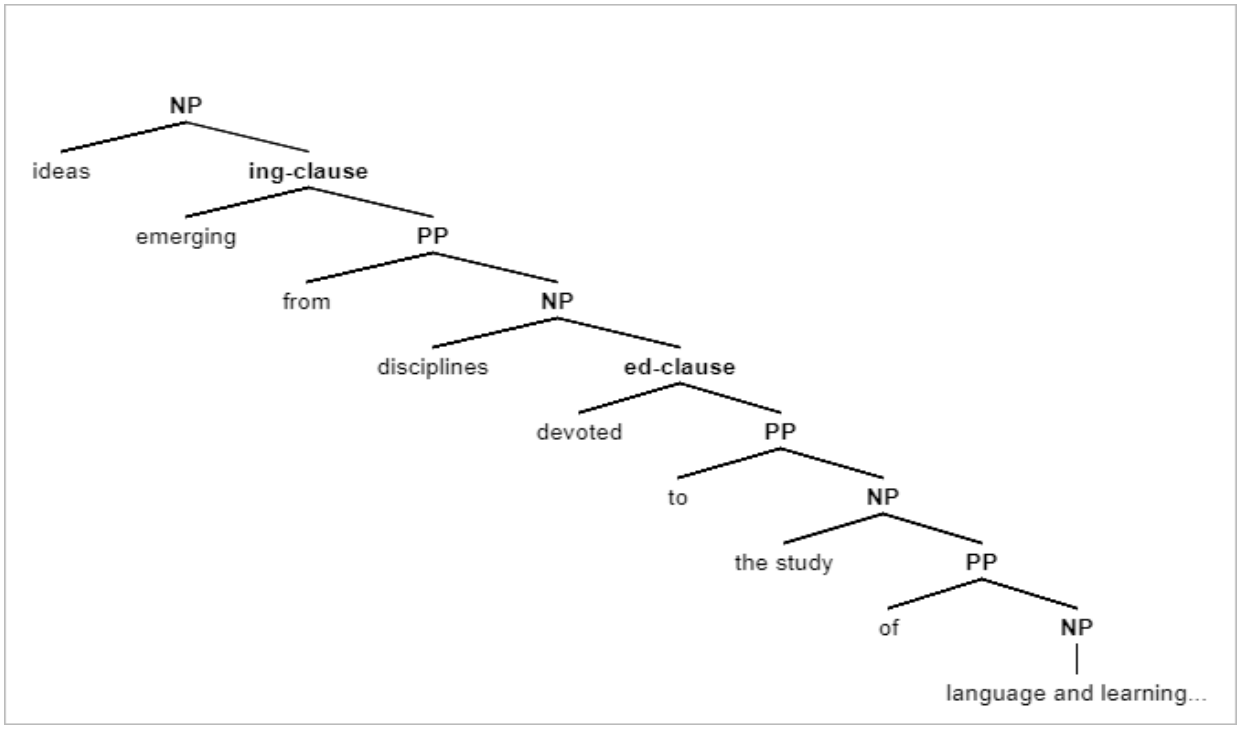
The complement–modifier dichotomy in these examples is, to quote Payne and Huddleston (2002:439–440), 'not as clearly differentiated syntactically'. Biber *et al*. (1999) also do not distinguish between modifying and complement functions of lexical and phrasal noun phrase elements and, given the fact that it is difficult to distinguish these functions syntactically, from here on out all phrasal constituents (nouns, adjectives and preposition phrases) functioning as dependents of the head noun will be referred to as modifiers.

In the case of postnoun clauses as constituents of the noun phrase, the functional distinction between complement and modifier can more readily be made on both semantic and syntactic grounds. Nouns can have postnoun modifiers or complements (or both in the case of stacking) or postmodifier complexes as illustrated in example (14) (Biber *et al*., 1999:576):

14. [**ideas** [emerging [from [**disciplines** [devoted [to [the **study** [of [language and learning]]]]]]]]]

Figure 2-1 below illustrates this same example in diagram format showing the various instances of embedded postmodifiers where each noun head has its own postmodifier, but the entire postmodifier, starting with 'emerging' and ending with 'learning', modifies the head noun 'ideas'.

Figure 2-1: Multiple embedding in the noun phrase



2.2.1 Prenominal modification

According to Biber *et al*. (1999:574, 588) and Payne and Huddleston (2002:444), nouns, adjectives and participial forms can all function as prehead modifiers in the noun phrase. According to Biber *et al*. (1999:589) general adjectives in attributive position are of the most common types of premodification used in noun phrases since they serve to characterise the qualities of nouns within a range of semantic domains such as size, age and colour. Attributive adjectives allow for the inclusion of additional information in the noun phrases both syntactically and semantically.

The following examples from Pullum and Huddleston (2002a:528) illustrate adjectives functioning as prehead internal dependents in the nominal (15–16). (The modifiers are underlined and the head noun bolded.) These are the prototypical instances of adjectival premodification relevant to this study:

15. my <u>new</u> **job**

16. <u>good</u> **work**

This naturally excludes adjectives in predicative (17) and postposed position (18), where the head referent is bolded and the modifiers underlined in each case (Biber *et al*., 1999:515, 519):

17. **she** seems quite nice really

18. **Something** cold and refreshing here

In the examples below, Biber *et al*. (1999:588) show how some attributive adjectives can be rephrased as a postmodifying clause, thus illustrating that the attributive adjective is the more dense choice. In (19) the adjective is in attributive position, whereas (20) contains the postmodifying option.

19. a big **pillow**

20. a **pillow** which is big

According to Biber *et al*. (1999:589), descriptive adjectives such as in the examples above are the most common category of premodifiers in all registers due to their wide semantic range as descriptors. In addition to central adjectives, the past participial and gerund participial (participial adjectives) are also part of the major structural types of premodification in English (Biber *et al*., 1999:588). Examples (21) and (22) taken from Biber *et al*. (1999:588) illustrate the *ed*-form in prenoun position and (23) and (24) of the *ing*-participial form:

21. a restricted **area**

22. established **tradition**

The *ing*-participial form has the same modifying function as the *ed*-forms above:

23. flashing **lights**

24. exhausting **task**

As with general adjectives, these noun premodifiers may also be rephrased into less dense cognates (Biber *et al*., 1999:588):

25. an **area** which is restricted

26. **lights** which are flashing

Based on these examples, Biber *et al*. (1999:588) claim that ‘premodifiers are consistently more condensed than postmodifiers, using fewer words (often a single word) to convey similar information’.

Although less frequent than central adjectives, nouns can also function as premodifiers of the

head noun in a noun phrase (Biber *et al.*, 1999:589). Noun + noun sequences can express a wide range of semantic relationships between the head noun and the modifier (Biber *et al.*, 1999:589). The absence of function words to express the logical relationship between the head and modifier results in a very dense packaging of information. Biber *et al.* (1999:588) supply the following examples of noun + noun sequences to illustrate the various meaning relationships between modifier noun and head noun and the various ways in which they can be rephrased:

27. plastic trays = trays made from plastic (not trays used for plastic)

28. pencil case = case used for pencils (not case made out of pencils)

The noun + noun sequence in (28) – *pencil case* – is denser than its syntactic counterpart, a *case used for pencils*, which incorporates a nonfinite clause as postmodifier. Biber *et al.* (1999:590–591) have shown that premodifiers aid in packing as much referential information as possible into a noun phrase. Regardless of the internal complexity of a noun phrase (noun heads with multiple premodifiers), Biber and Gray (2011) argue that premodification is structurally more economic or denser than using postmodifying and complementation devices.

2.2.2 Postnominal modification

The post-head position in the noun phrase is reserved for a range of structures fulfilling the function of (post)modifier or complement. An exposition of the distinction between modifier and complement will be given in the section on complements later on. This section will focus on the formal types of postmodifiers. Both phrasal and clausal structures, that is, constituents in the form of phrases and constituents in the form of clauses, have a role in noun phrase modification (Biber *et al*., 1999:575).

The major structural types of postmodification are preposition phrases, relative clauses and nonfinite clauses (participial clauses and *to*-infinitive clauses) (Biber *et al*. 1999:604). Biber and Gray (2012:323–326) show that these structures are also responsible for creating syntactic constructions with varying degrees of density. Biber *et al*. (1999:634) point out that some postmodifying preposition phrases can also be rephrased as full relative clauses in the same way the premodifiers have postmodifying counterparts. To briefly illustrate the structural difference between a postmodifying phrase and postmodifying clause, consider the following examples from Biber *et al*. (1999:575):

29. **compensation** <u>for emotional damage</u> [prepositional phrase]

30. the imperious **man** <u>standing under the lamppost</u> [nonfinite clause]

31. enough **money** to buy proper food [nonfinite clause]

In (29), the postmodifier *for emotional damage* is in the form of a preposition phrase, and in (30) *standing under the lamppost* and (31) *to buy proper food* both are clauses filling the postmodifier function. Where phrasal modification is understood to refer to modification through the means of phrasal structures, clausal modification has to do with clausal structures responsible for modification. Biber *et al*. (1999:575) and Payne and Huddleston (2002:445) list additional forms of postmodifier, including appositive noun phrases, adverbs, adjectives and nouns. However, these postmodifiers were not specifically analysed in this study, given that they are much less common than prepositional phrases and clauses as postmodifiers.

2.2.2.1 Postnominal phrasal modification

According to Biber *et al.* (1999:604–606) prepositional phrases are the most common device used for modification within the noun phrase, making up the majority of postmodifiers (including relative clauses) across all registers, but especially in the expository registers. Further, Biber and Clark (2002:58–59) have shown that preposition phrases far outweigh the frequencies of postmodifying finite and nonfinite clauses (clausal postmodifiers) not only across different registers, but also over time, where preposition phrases, *of*-phrases in particular, have significantly increased from the eighteenth to the twentieth century.

Biber *et al*. (1999:635) stipulate that *of*, *in*, *for*, *on*, *to*, *with*, *about*, *at*, *between*, *by*, *from* and *like* are the twelve most common prepositions in the *LGSWE*. Based on the frequency rate of these structures, these prepositions were included for this analysis. The postmodifying function is illustrated in the following examples from Biber *et al*. (1999:635–637):

32. ten **words** of English

33. the **mess** in his bedroom

34. a **school** for disabled children

35. the **search** for new solutions

36. his most wounding **attack** on the tabloids

37. their first **trip** to Scotland

38. the **man** with the megaphone

39. a sensible **relationship** with the West German mark

As mentioned earlier, this study will follow Biber *et al.* (1999) in not distinguishing between modifying and complementation roles of postnoun prepositional phrases as Pullum and Huddleston (2002b:661) do in the quantitative analysis of constituents of noun phrases.

2.2.2.2 Postnominal clausal modification

According to Huddleston, Pullum and Peterson (2002:1034), relative clauses are a type of modification device where the relative structure is directly related to an antecedent to which it refers. Biber *et al*. (1999:608) emphasise that these modifiers specifically contain information pertaining to the head referent, which is usually a person or thing. Relative clauses may be introduced by the relative pronouns *which*, *who*, *whose*, *whom* and *that* and the relative adverbs *when*, *where* and *why*, where the interpretation of the relativiser is determined by the antecedent (Huddleston, Pullum & Peterson, 2002:1034).

The following examples illustrate the use of the various relativising pronouns for postmodification. These examples are taken from Biber *et al*. (1999:608):

40. The lowest pressure **ratio** *which* will give an acceptable performance is always chosen

41. There are plenty of existing **owners** *who* are already keen to make the move

42. There was a slight, furtive **boy** *whom* no one knew

43. It was good for the **fans**, *whose* support so far this season has been fantastic

44. Well, that's the only **way** *that* this can be assessed

Huddleston, Pullum and Peterson (2002:1034–1039) provide similar examples of noun relativisation:

45. They ignored the **suggestion** *that* Kim made

46. Focus on the **question** *which* your brother raised

47. The **boys** *who* defaced the statue were expelled

48. We've never met the **people** *whose* house we are renting

49. the **area** *where* the chapels have closed

50. one **day** *when* she was at school

51. the other **reason** *why* the ambulance workers have lost out

The examples from Huddleston, Pullum and Peterson (2002:1045) express these same meaning relationships:

52. they want to go to the **place** *where* they went last year

53. it was a **time** in my life *when* everything seemed to be going right

According to Biber (2010:167), relative clauses are a salient feature of information-dense discourse and are remarkably less common in informal registers such as conversation. Relative clauses functioning as noun postmodifiers introduce clausal material into the noun phrase without expanding on the verb of the main clause itself.

Biber *et al*. (1999:602) further distinguish between restrictive and nonrestrictive relatives, and claim that all postmodifiers have either a restrictive or nonrestrictive function, whereas Huddleston, Pullum and Peterson (2002:1034) distinguish between integrated and supplementary relatives. They define restrictive or integrated relative constructions as:

> [relatives that are] integrated into the construction containing them, both prosodically and in terms of their informational content. The prototypical integrated relative serves to restrict the denotation of the head nominal it modifies, and is often referred to by the term 'restrictive relative' (Huddleston, Pullum & Peterson, 2002:1034–1035).

Their definition of a supplementary construction reads:

> A supplementary relative clause adds extra information about the antecedent, information not fully integrated into the structure of the containing clause and not needed to delimit the set denoted by the antecedent (Huddleston, Pullum & Peterson, 2002:1035).

This distinction is illustrated by Huddleston, Pullum and Peterson (2002:1034) in (54) and (55):

54. The **boys** *who* defaced the statue were expelled [integrated]

55. My **father**, *who* retired last year, now lives in Florida [supplementary]

In (54), the relative clause clearly specifies which *boys*, but in (55) it provides additional information with reference to the head of the phrase, which is in fact the function of modifiers. Integrated relative constructions align with Biber *et al*.'s (1999:195) restrictive relative clauses, while supplementary relative constructions align with nonrestrictive relative clauses. Biber *et al*. (1999:605) continue to categorise both restrictive and nonrestrictive functions as modifiers, and since postmodifying clauses serve not only to identify the referent but also to 'add some descriptive information about the noun' (Biber *et al*., 1999:645), no distinction was made in the

data analyses between the two subfunctions.

Nonfinite clausal structures are another category that can function as noun postmodifiers (Biber *et al*.,1999:630; Huddleston, 2002:1264). Nonfinite clauses are the participial clauses which include the gerund participial (*ing*-clauses and past participial (*ed*/*en*-clauses), and the *to*-infinitive clauses (Biber *et al.*, 1999:630; Huddleston, 2002:1264). To briefly illustrate the difference between finite and nonfinite relatives, consider the examples from Biber *et al*. (1999:630) below, where (56) contains a nonfinite participial form and (57) the corresponding finite relative clause. In both, again, the head noun is in bold and the modifier underlined:

56. **selections** retained from the second year

57. **selections** which are retained from the second year

Gray (2015:65–66) argues that nonfinite clauses are denser than their full relative alternatives. Biber *et al*. (1999:631–632) makes the same observation. In the case of *ed*-clauses, both forms – nonfinites and full relatives – are available and interchangeable for use. Therefore, the preference for *ed*-clauses over full relatives is a matter of economy, where the same information is conveyed through fewer words (Biber *et al*., 1999:632). As for *ing*-clauses, the verb often places limits on whether a full relative is a viable option or not, and therefore structural considerations limit choice (Biber *et al*., 1999:632).

The gerund participial *ing*-clauses (58) and past participial *ed*-clauses (59–60) in postmodifying function taken from Biber *et al*. (1999:630–631) are illustrated below:

58. Interest is now developing in a theoretical **approach** involving reflection of Alfvén waves

59. The US yesterday welcomed a **proposal** made by the presidents of Colombia, Peru and Bolivia

60. It can be derived using the **assumptions** given above

Huddleston's examples are similar (2002:1264):

61. **People** living near the site will have to be evacuated

62. I came across a **letter** written by my great-grandfather

The *to*-infinitive clause in a noun phrase can perform the same postmodifying function as *ing*- and *ed*-participial clauses (Biber *et al*., 1999:632). Consider the following examples from Biber *et al*.

(1999:632–634):

63. They'd take a long **time** to dry

64. There is one further **matter** to confess

65. She's had a **lot** to put up with

This section served to provide an overall outline of modifying constructions in the noun phrase, and although not all types of modifiers were included in the study, the most frequent ones selected for this study should be sufficient for comparing registers with regard to densification strategies. While both modification and complementation are strategies to alter the meaning of the head noun, complement clauses are structurally different from modifying clauses. The section below discusses clauses that have a complementising rather than modifying function within the noun phrase.

2.2.3 Postnominal clausal complementation

This final section considers the role of complementation in the noun phrase. Two types of noun complementation are identified by Biber *et al*. (1999:645). These are *that*-complement clauses and *to*-complement clauses. According to Biber *et al.* (1999:604), these clauses are structurally and semantically different from the postmodifying function. The difference between *that*-clauses fulfilling a postmodifying function and those fulfilling a complementising function is illustrated by the examples from Biber *et al*. (1999:644):

66. Peter reached out for the well-thumbed **report** that lay behind him [modifier]

67. Other semi-conductor stocks eased following an industry trade group's **report** that its leading indicator fell in September [complement]

Further, *that*-complement and *to*-complement clauses are governed by a select group of nouns, and these nouns form an abstract relationship with the complement clause (Biber *et al*., 1999:647) Biber *et al*. (1999:648; 653) use the following examples to show the relationship between the head noun and the complement:

68. But there remained the very troublesome **fact** that leguminous crops required no nitrogenous manure

69. There is a **possibility** that this morphology represents an ancestral great ape character

70. Their frustrations were the product of their **belief** that the leadership was not responding adequately

These same relationships are also expressed by *to*-complement clauses:

71. We need to give people a decent **chance** to elect a sensible council

72. Last year the society's committee made a **decision** to relaunch in a **bid** to attract more members

Roughly speaking, the relationship between the head noun and the complement can be characterised as being equal where the complement is the referent; for example, *that the leadership was not responding adequately* is the *belief* (70), *to elect a sensible council* is the *chance* (71) and *to attract more members* is the *bid* (72).

As far as their distribution across registers goes, Biber *et al*. (1999:647) show that *that*-complement clauses are also more frequent in academic writing, whereas *to*-complement clauses are especially prevalent in news; both structures are the least common in conversation.

Finally, Biber *et al*. (1999:655–656) include the use of nouns followed by an *of*- plus an *ing*-clause (73) and nouns followed by *wh*-clauses (74) as complementisers. For instance:

73. Feynman discusses the **idea** of putting a lamp between the two slits to illuminate the electrons

74. We have no **knowledge** of where it came from

The relationship between the head and the clause is similar to that of *to*-complements and *that*-complements.

2.3 Conclusion

Grammatical features creating noun phrase density are the central structures of this analysis. As the grammatical framework, Chapter 2 provided a general overview of the noun phrase structure, that is, the immediate elements involved in modifying and complementising the head noun. The examples provided here from the reference grammars form the base for creating the relevant search strings used in the corpus data. Since the syntactic structures can also fill a variety of other functions not related to density, such as complements of the verb or adjective, or as adverbial in the clause, these functions are weeded out of the concordance data in the coding process. Therefore, semantic distinctions are important to the coding process and analysis. Chapter 3 relays the methods involved in arriving at a workable dataset for the analysis.

CHAPTER 3 METHODOLOGY

3.1 Introduction

The aim of this study is to investigate how noun phrase density is realised in web-based registers, compared to offline registers. In order to do this, these features have to be identified before they can be compared and described in terms of their similarities and differences in terms of their distribution patterns. This chapter sets out the method in answering this question, by discussing the data, specifically the collection methods and statistical measures used in selecting and sampling the concordances for each nominal modification pattern. It emphasises the role of basic descriptive statistics such as frequency counts, and extends the use of keyness analysis techniques, beyond lexical comparisons, to measure the differences between grammatical constructions. The first section outlines the architecture of the two main corpora and the subregisters selected within those two sets for comparison. The second section describes the role of basic descriptive statistics and its application to the data. This includes the formation of search strings to identify grammatical patterns in the data, the extraction of frequency data in the subregisters, hand-coding of concordance samples and the creation of raw frequency estimates used for presenting normalised data. Further, this chapter touches on the role of log likelihood as a measure of significance, but more importantly, it emphasises the use of effect size measures in identifying important differences between the subregisters. This study therefore largely draws on keyness analysis methods to make 45 pairwise comparisons. The size of the difference between constructions is measured using a percentage difference metric (Gabrielatos & Marchi, 2012). Each group of patterns can then be ranked according to the size of the difference, making it possible to establish which constructions and which registers are frequent contributors in creating larger and smaller differences.

3.2 Corpora

The corpora chosen for this study were selected based on considerations relating to structure and composition that allow the exploration of nominal group structure across registers. In order to do this, at least one very specific type of corpus was needed, namely a corpus constructed from texts originating on the internet. This is the Corpus of Online English, henceforth referred to as CORE (Davies, 2016-). The reference corpus is the Corpus of Contemporary American English, or COCA (Davies, 2008-).

CORE is the product of the originally compiled Global Web-based English or GloWbE corpus (Davies, 2013). GloWbE is a 1.9-billion-word corpus consisting of texts from online webpages compiled from the year 2012 to 2013 (however, actual texts may date back further). The

webpages range from more formal texts (40%), including formal emails and news articles, while the other 60% comprises informal genres – mostly blogs (Davies & Fuchs, 2015:2–3). CORE is constructed from the online texts in GloWbE, but with the structural design that makes CORE ideal for investigating various registers on the web and therefore for answering this study's research question.

Biber, Egbert and Davies (2015) created a taxonomy that end users used to classify the CORE texts into subregisters. The coding process yielded 33 different registers in CORE, and the entire corpus contains just over 50 million words. The categorisation is viewed as reliable and balanced.

Out of the 33 registers, only five were selected for the analysis. These are Interactive discussion, Personal blog, News, Historical article and Frequently asked questions (FAQ). Some registers such as Lyrical, TV/movie script, Research article and Short story were excluded from the analyses, because many of the texts within these registers are uploaded copies from original offline versions. The metadata that contains the URL links for each webpage were randomly searched in order to establish whether the page originated on the web or not. If the links showed a general tendency toward a web-born status, these were included. There are of course online-based registers in CORE that were not included. A reasonable selection of registers ranging from popular to those with more informational purposes were selected based on practical considerations relating to the scope of the investigation.

COCA is 450-million-word corpus that is updated regularly and is specifically constructed for register comparison purposes. This corpus contains five offline registers that were selected for the comparative study. They are Spoken, Fiction, Magazine, News and Academic. These offline texts serve as the nonweb counterpart for the registers selected in CORE. Additionally, COCA allows for diachronic investigations, while CORE does not. However, both corpora fit the contemporary English category and therefore these corpora are readily comparable.

3.2.1 CORE subregisters

The CORE registers are an online (web-born) medium in the written mode. CORE registers all have the option of being revised and are not viewed as 'produced in real time' in the same way that a conversation would be, despite some being more oral-based and involved than others. There is no immediate shared context, but authors may assume shared context despite not being familiar with the reader or audience. The communication ultimately takes place in a public domain, despite being produced without an immediate audience.

As already mentioned, the online registers selected for this study are Interactive discussion, Personal blog, Historical article, News and FAQ, and they form part of CORE. These registers

have varying communicative purposes. Interactive discussion is marked for an oral-involved and elaborative style (Biber & Egbert, 2016:108–111), mimicking real-time speech between multiple participants but in written form. Personal blog has been generally categorised as having a narrative–communicative purpose. Despite its written mode, it is marked with features associated with oral involvement (Dimension 1) and oral elaboration (Dimension 2) as opposed to informational purpose (Biber & Egbert 2016:108–111).

Historical article has both an information-driven communicative purpose as well as that of narration. It has a high frequency of features associated with noun modification on both a literate–informational and informational narration dimensions, but without an involved narrative style (Biber & Egbert, 2016:108–111). Online News is generally classified as narrative in style, but is also marked for nominal and noun modification features on dimensions characterised with functional purposes of conveying and reporting on information (Biber & Egbert, 2016:108).

Frequently asked questions about information are only marked for nominal features on one dimension labelled nontechnical description by Biber and Egbert (2016:108–111). In terms of communicative purpose, FAQ seeks to describe and explain information, is procedural in nature and does not convey information in a narrative style in the way that Personal blog, online News and Historical article do.

3.2.2 COCA subregisters

The five offline registers are from COCA and include Academic as the most information-driven component, News, Magazine and Fiction as more popular writing and finally Spoken. Spoken is largely made up of 'unscripted' transcribed speech from various media programmes. This register also includes scripted speech from news programmes, which introduce additional communicative purposes such as reporting or describing events. Spoken contains transcripts of unscripted conversation from different TV and radio programmes such as *NewsHour* (PBS), *Good Morning America* (ABC), *Today Show* (NBC) and *Jerry Springer* (Brigham Young University Corpora Website [BYU]). This register can be viewed as produced under real-time constraints. The discourse style is heavily involved and interactive with multiple participants and is accessible (despite the topic) to a wide or popular audience.

Fiction is made up of short stories and plays from literary magazines, children's magazines, popular magazines, first chapters of first edition books (1990 to present) and movie scripts (Brigham Young University Corpora website [BYU]). It contains text designed to provide entertainment through narration and, like most written registers, is uninvolved. It is aimed at a wide and general readership.

Magazine includes a mix of edited texts related to various topics, including news, health, home and gardening, women, financial, religion and sports (BYU). These texts are noninteractive and accessible to a general audience, and the content includes communicative purposes such as expressing opinions, giving advice, and narrating and describing things and events. Therefore, the communicative purposes may be mixed, given the array in topics and texts included.

News, being noninteractive, is largely aimed at reporting and narrating events and undergoes extensive revisions before being published. This register contains news articles from *USA Today*, *New York Times*, *Atlanta Journal Constitution* and the *San Francisco Chronicle*, but also varies in reported content such as finance and sport (BYU). The readership is more general, but it probably depends on the topic. The writer's perceived knowledge of the readership's frame of reference will affect the writing style of news articles.

Academic consists of an array of scientific topics from peer-reviewed journals, including world history, technology and psychology (BYU). This subregister generally aims to describe information objectively in a formal discourse style. These texts are aimed at a more topic-specific audience and are therefore not involved or interactive. Academic writing also undergoes extensive revision. Table 3-1 summarises the main characteristics of the subregisters involved. This table also informs the interpretation of the analyses.

Table 3-1: Situational characteristics of registers in CORE and COCA

Register	Production circumstances	Communicative purpose	Setting	Participants	Channel (mode & medium)
Interactive discussion	Option of being revised/edited 'more' real time	To interact/discuss	The author assumes shared knowledge / immediate	Multiple addressors Wide audience	Writing – web
Personal blog	Option of being revised/edited	To narrate	Public domain / not shared / author assumes shared knowledge	Single/co-authored Wide audience	Writing – web
Historical article	Revised/edited	To narrate	Public domain / not shared / author assumes shared knowledge	Single/co-authored Wide audience	Writing – web
News	Revised/edited	To narrate	Public domain / not shared / author assumes shared knowledge	Single/co-authored Wide audience	Writing – web
FAQ about information	Revised/edited	To describe/explain information	Public domain / not shared	Single/co-authored Specialist audience	Writing – web
Spoken	Real time (possibly scripted)	Mixed – to report/ entertain/describe	Public / shared vs nonshared	Multiple participants	Transient speech (transcribed)
Fiction	Revised/edited	To narrate	Public domain / not shared	Single/co-authored Wide audience	Writing – offline (typed)
Magazine	Revised/edited	Mixed – to report/ entertain/describe/ advise	Public domain / not shared	Single/co-authored Wide audience	Writing – offline (typed)
News	Revised/edited	To narrate/report/inform	Public domain / not shared	Single/co-authored Wide audience	Writing – offline (typed)
Academic	Revised/edited	To describe/exposit	Public domain / not shared	Single/co-authored Specialist audience	Writing – offline (typed)

Given what previous studies have shown with regard to noun phrase modification and register variation, the expectation is that expository writing and informational–descriptive texts are more productive in their use of nominal modification strategies. The categorisation and multidimensional classification of the CORE registers (Biber & Egbert, 2016) suggest that Interactive discussion and Personal blog will have the lowest frequencies of modifiers, while Historical article, News and FAQ have the higher scores. Similar predictions hold for Spoken, with the lowest uses, which increase along the cline of Fiction, Magazine, News and Academic. While online registers have already been classified, little is said about how noun modification marks differences in relation to other offline written and spoken registers. There may be differences between pragmatically similar registers with regard to modifier type.

3.3 Concordance data

In comparative corpus linguistic studies, frequency counts are a very basic statistic used to provide descriptions for the occurrences of linguistic phenomena (McEnery & Hardie, 2012:49; McEnery & Wilson, 2001:82). Frequencies are indicative of the extent to which a particular linguistic feature is present in any corpus, but in their raw form, frequency counts reveal little about the prevalence of a feature in terms of its proportion in one corpus compared to another (McEnery & Wilson, 2001:82). In order to make corpora comparable in terms of sheer numbers, raw frequencies are converted through normalisation. Calculating frequencies as a proportion of the corpus size creates an equal basis for comparison of the features between differently sized corpora. Normalised frequencies are also the values used in some effect size metrics (Gabrielatos, 2018:235), which are discussed later in this section.

The frequency data for modification patterns in the subregisters of COCA and CORE were obtained either by means of the BYU corpus online interface[1] or through a specifically compiled script[2] that successfully counted patterns, derived a count that formed the basis of normalisation (discussed further below) and produced concordances for each pattern in each individual register. When searching linguistic patterns using the online interface, frequency counts were displayed per pattern per register per year or group of years, depending on the corpus structure. In CORE, individual frequencies for every search string were recorded as the total number of hits per register. With regard to COCA, frequencies for premodification patterns were obtained via the script, also arriving at a total number of hits per pattern per register. Concordances relating to postmodification in COCA could be obtained using the online interface. Each register under investigation here therefore has similar structural parameters.

3.3.1 Search strings

Both CORE and COCA are also part-of-speech tagged by the CLAWS7 tag set, making it possible to search for words tagged as nouns and the relevant modifying structure either preceding or following the head noun[3].

With a common noun as the head, the prenominal modifiers identified were common nouns (1),

---

[1] The Brigham Young University Corpora website hosts both COCA and CORE (and other corpora) in an online format and is available at https://corpus.byu.edu/overview.asp.

[2] The script was compiled to replicate the search strings and output in offline data, which were largely inaccessible on the BYU interface.

[3] For postmodification, _n* was used to identify the head noun and would therefore include common nouns and proper nouns in the head position.

proper nouns (2) and adjectives (3)[4].

1. The departure lounge is large (Interactive discussion)

2. She is passionate about the Oxford comma (Personal blog)

3. the sheer waste of expensive ammunition (Historical article)

The examples above are distinguished from instances where the head noun takes the form of a common noun but functions as a proper noun (4), or where a common noun precedes another common noun but the unit as a whole functions as a proper noun (5).

4. John Diamond (Academic)

5. World Trade Center (Spoken)

Postmodifying prepositional phrases were identified through searching for any noun followed by a preposition[5]. These included *of*, *in* and *for*, along with *on*, *to* and *with*, which account for 90% of all prepositional phrases as postmodifiers, in the *LGSWE*, and the remaining 1%, that includes *about*, *at*, *between*, *by*, *from* and *like* (Biber *et al*., 1999:635). Concordance lines for all twelve prepositions were evaluated in terms of its postmodifying function. Structurally, the following grammatical options (6-7) are viable when searching for nouns followed by prepositions. Example (6) is the only phrasal postmodifier, example (7)[6] is a tag error and the structure in (8) fills the role of adverbial:

6. a regular rider of the MetroLink system (COCA News)

7. parents have to option to opt in (Spoken)

8. demonstrate this knowledge in your cover letter (Magazine)

Additionally, the _n* [preposition] + _ii* string would yield complement *ing*-clauses (9) and prepositional relative clauses (10). These instances were coded separately and included as part

[4] _nn* + _nn* was used for noun + noun sequences, _np* + _nn* for proper noun as premodifiers and _jj* + _nn* for adjective noun sequences.

[5] The 12 most common prepositions identified by Biber *et al*. (1999) were searched individually by _n* + [preposition]_ii*. The _ii tag identifies general prepositions. For, of and with each have a unique tag. Noun + *for* patterns are identified by _n* + *for*_if*, noun + *of* was identified by _n* + *of*_io* and noun + *with* sequences were found using _n* + *with*_iw*. The relevant preposition tag was included per prepositional phrase search string.

[6] The *to*-clause to opt in was yielded in a *to*-preposition phrase concordance sample,

of the count for prepositional clausal postmodifiers, distinguishing them from phrasal postmodifiers (and postmodifying relative clauses without prepositions):

9. A preliminary work model with an emphasis on counseling professionals on university campuses (Academic) [preposition + complement *ing*-clause]

10. the desert in which she was walking (Magazine) [prepositional relative clause]

Postmodifying relative clauses were identified similarly to phrasal postmodifiers, but with the introducing relativiser following the head noun[7]. In other words, the pattern is noun + *that* or noun + *which*. Except in the case of adverbs *where*, *when* and *why*, which were grouped as noun + *wh**, individual strings were constructed for each relativiser. *That*-relatives, as well as the relative pronouns, *which*, *who*, *whose* and *whom*, were searched as individual patterns, each having its own tag. Concordance samples, again, were hand-coded to separate nominal postmodifying relative clauses (examples 11–17) from other syntactic functions, such as complement (15), extraposed position (16) and clause-level complement (17). *That*-complement clauses were therefore identified within the same concordance as that of *that*-relative clauses:

11. He had brown **hair** that was shaggy near his collar (Fiction) [relative clause]

12. the **trophies** which vain mortals have by wit, or valor, or by virtue won (Academic) [relative clause]

13. the man is also a **thief** who has stolen a million dollars (COCA News) [relative clause]

14. the **situation** where you can't even have, you know, use a Bluetooth (Spoken) [relative clause]

15. their **prognosis** that Gerald would never again be himself (Fiction) [*that*-complement]

16. it's a **no-brainer** that he should be in the Hall of Fame (COCA News) [extraposed *that*-clause]

17. promising not to tell her **husband** that the kids did it (Magazine) [clause-level complement; direct object]

[7] Search strings for marked relatives are _n* + [relativiser]. Relativiser that has a unique CST tag, distinguishing it from other pronouns and determiner that.

The estimated frequencies for unmarked relative clauses (zero relatives) are based on a pattern that incorporated the use of a personal pronoun[8] instead of a noun as head. Unmarked relative clauses do not have an overt grammatical unit identifiable by a POS-tagger and therefore are not completely accounted for in the data. Using the pronoun as head (instead of a noun) offered an alternative way to account for these structures to some extent, and would yield examples such as (18) and (19):

18. all the **things** we do not have (Personal blog)

19. the few **men** he had managed to collect together (Historical article)

Unmarked relative clauses where the head noun was followed by another noun as in (20) would therefore have been excluded from the analyses:

20. all the **things** people do not have

Finally, participial nominal postmodifiers, that is, noun + *ed*/*en*-clauses, noun + *ing*-clauses and noun + *to*-clauses[9] were searched using tags from the online interface. The tags _v?n* and _v?g* identify clauses introduced by nonfinite verbs ending in *ed/en* (21) and *ing* (22) and would result in constructions such as:

21. a **flagpole** topped with a liberty cap (Historical article) [*ed*/*en*-clause]

22. You may be asked to join a **study** looking at side effects (FAQ) [*ing*-clause]

Instances of tag errors in example (23) and other syntactic functions such as an adverbial on clause level (24) and object predicative in the case of complex transitive constructions (25) were eliminated through the manual coding process:

23. British **negotiators** advocated for a native buffer state (Historical article) [tag error]

24. I had some bad **results** using this (Interactive discussion) [clause-level adverbial]

25. labour movement defeats left **workers** looking for an alternative (Historical article) [object predicative]

Concordances containing nonfinite to-infinitive clauses were retrieved by searching for any noun

[8] Strings for unmarked relatives are _nn* + pronoun + verb.

[9] As with phrasal and finite clausal postmodifiers, search strings for participle clauses were structured individually and takes _n* as the head in the string.

followed by the infinitive marker to, which is has a _TO[10] tag. This concordance yielded *to*-postmodifying clauses and *to*-complement clauses respectively. Each concordance sample for this structure was coded to separate these two functions. Searching for *to*-infinitives tagged as _TO separated cases of *to*-complementation (26) from *to*-clause postmodifiers (27), and the coding process would ultimately exclude tag errors (28) and syntactic functions such as direct object (29) and object predicative (30) from the concordance data:

26. the **decision** to run for president (Spoken) [complement]

27. our list of **places** to visit (Personal blog) [modifier]

28. something different **needs** to happen (Academic) [tag error]

29. ask **students** to select ten key events (Academic) [direct object]

30. water causes **salt** to dissolve (Academic) [object predicative]

The modifier complement structures identified in this section are the density features investigated in this study.

3.3.2 Precision rate and estimated frequency

The number of hits for any given search string represents surface structure only. To arrive at a true frequency for each relevant density device, the precision for each concordance was established manually coding two random samplings from each concordance to identify how many of the hits were in fact instances of the modifying and complementising functions within the noun phrase.

The manual coding process involved analysing a sample size of 400 lines (two samples of 200 lines each) per register per pattern; where the total number of hits were 2000 or below per structure, two samples of 100 concordance lines were analysed. In exceptional cases where the total number of hits were extremely low (500 or below), the entire sample was coded. The total number of surface structures (search strings) amounted to 25 patterns. Approximately 83 971 concordance lines making up this data-set were analysed. Of the 25 patterns, 39 density features ultimately make up the data-set. For instance, a search string for *that*-relative clauses would also retrieve hits for *that*-complement clauses and therefore the search string captures two features under one external structure. Of the 39 features, 0.20% of features in CORE and 0.07% in COCA

[10] The _TO tag is unique to the infinitive marker *to* in the CLAWS7 tagset and should ideally result in structures containing *to*-clauses rather than *to*-phrases.

exceeded a 7% difference in range. The majority of concordance samples for each feature were therefore consistently coded and relatively similar.

Once all the concordances were hand-coded to improve the precision rate for each density feature, the final frequency score was captured in an Excel spreadsheet for each register. This sheet was used for further percentage and estimate calculations. For each individual feature, the number of constructions were divided by the sample size in order to arrive at a percentage of combined frequencies across the two samples. This percentage was then multiplied by the total number of original hits of the surface pattern, finally arriving at an estimated frequency per feature.

Table 3-2 illustrates the calculation performed on frequencies of proper noun + noun patterns in Fiction. The first column contains the number of raw hits for proper noun + noun sequences. Columns 2 and 4 contain the frequencies of the type of construction as two individual concordances. Column 3 and Column 5 indicate sample size and Column 6 is the average of combined frequencies between two samples. The calculated frequency in Column 7 is the result of the average multiplied by total number of hits.

Table 3-2: Estimated frequency calculation for nominal modification and complementation

Fiction (COCA)						
Total number of hits	**Number of constructions in Sample 1**	**Sample size 1**	**Number of constructions in Sample 2**	**Sample size 2**	**% of constructions combined samples**	**Calculated estimated frequency**
36 749	56	200	55	200	0,2775	10 198

This estimate (the value in the last column) better reflects the frequencies for each pattern that would have been arrived at had it been possible to reliably search for distinctions within functions. The estimates are therefore those which are used to compare each of the constructions across the selected registers. These estimates now form the base of comparison and are normalised to a base of 100 000 nouns (and for comparative purposes 100 000 tokens). Normalised frequencies also inform keyness and effect size measures, which are discussed later in this chapter.

3.3.3 Relative frequency and noun frequency as basis of normalisation

It was previously mentioned that normalisation is a calculation of the percentage by which a grammatical item occurs in a corpus. It establishes how often a feature, typically a lexical item, appears per x amount of words or tokens and is expressed as 'so many times per 1 000 or per 100 000 words' (the normalisation base depends largely on corpus size and the relative frequency of a phenomenon, so as to avoid a normalised number that is below 1) (McEnery & Hardie, 2012:49–50). However, Brezina (2018:104) emphasises the importance of considering the linguistic context when investigating occurrences of grammatical features in different texts. He refers to this context as the lexicogrammatical frame, which restricts the scope in which particular linguistic variables will appear (Brezina, 2018:107).

In this case, the linguistic context concerns nominal modification, and therefore the total number of nouns occurring in each subregister forms the basis of normalisation in these analyses. Nominal modification is context sensitive and only possible in the context of a noun-headed noun phrase. Therefore, using the number of nouns as the devisor would be more accurate in describing the patterns of noun modification across registers. For the purposes of contrasting the differences in pattern variation when alternating the normalisation base, the same estimated raw frequencies were normalised using the word count of each subregister.

Figure 3-1: Noun count versus token count as basis for normalisation

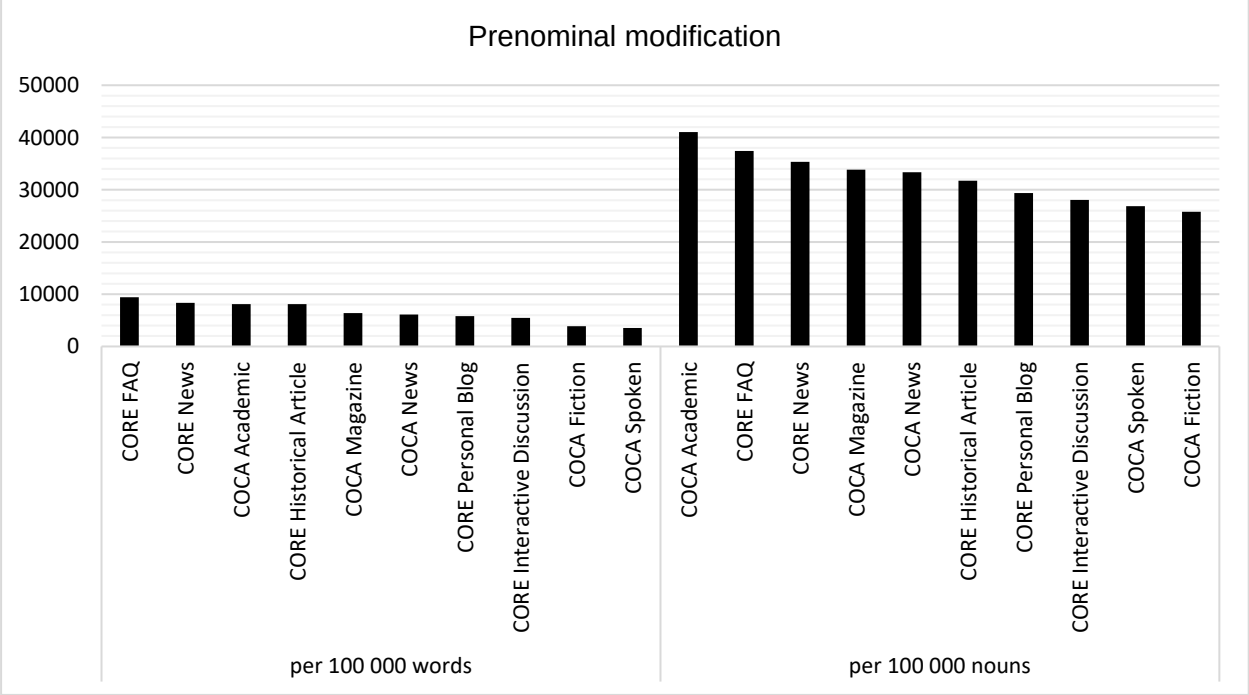


Figure 3-1 shows that when the base for normalisation is per total number of words (tokens), a different picture emerges. What the word-based data show is merely the number of times a

particular pattern occurred. For instance, the group of registers on the left of the graph shows that modifier + noun sequences are more common in FAQ and the least common in Spoken. This says nothing about the proportion of nouns modified in one register in relation to another.

The noun-based version, on the other hand, makes a claim regarding the likelihood of modification in each register, that is, if a noun occurred it is more likely to be modified in Academic, for instance, than in any of the other registers.

So, where general lexical items are compared across corpora, the use of token-based normalisation is appropriate. But where grammatical categories (the occurrence of which is restricted in terms of linguistic context) are compared across corpora, a narrower base for normalisation is necessary. In light of this, the frequency data here is normalised to 100 000 nouns and tokens, since the proportion of density features between online and offline registers are mainly the point of interest in the analysis. The following section further describes the pairwise comparisons for each of the density features.

3.4 Statistical pairwise comparison of registers

The normed frequencies form the basis of the next step in the analyses. Total noun counts were recorded for each subregister in COCA and CORE[11] and captured, along with the observed frequencies, onto an Excel spreadsheet[12]. This resulted in 45 pairwise comparisons. While it is now possible to compare the relative frequencies of each construction across registers, descriptive statistics are needed to report the extent to which registers differ or simulate one another. These measures are built into the spreadsheet.

The Excel sheet contains formulae for calculating a log likelihood value, where the *p*-value is then derived, and implements a variety of effect size metrics. One such measure is percentage difference (%DIFF), which is especially advocated by Gabrielatos (2018) and Gabrielatos and Marchi (2012). This metric and its application and log likelihood are discussed further below. Each pairwise comparison now contains information relating to the level of statistical significance and the meaningfulness of the differences observed, which forms the basis of the qualitative description of the data.

[11] The total number of common nouns (_nn*) in COCA 2008–2012 was counted with a script. The total number of nouns (_n*) in both COCA and CORE used for postmodification was obtained from the BYU online interface. Searching for _nn* patterns in the online interface proved problematic, since there were too many hits to display.

[12] The UCREL log likelihood and effect size calculator is available at http://ucrel.lancs.ac.uk/llwizard.html. The Excel spreadsheet was created by Paul Rayson and is implemented on this same page.

Keyness measures, a longstanding basic technique used in linguistic comparisons, are employed to identify those items, usually words, occurring significantly higher or lower in a target text when compared to a reference corpus (Pojanapunya & Watson Todd, 2018:133). These measures are usually built into existing linguistic analysis software (online significance test calculators, WordSmith Tools and so on) and allow the identification and further analysis of those important items that convey something about the style of the study text in question. In this case, the key items (nominal modifiers) were selected prior to the investigation of which the estimated frequencies formed the basis of comparison between the 10 subregisters discussed earlier in this chapter. In other words, these specific constructions will be key for certain text types (registers), but the extent to which they are key for these particular registers is unknown and ultimately where significance testing and effect size statistics play a role.

3.4.1 Significance tests

Depending on the type of research question, linguistic inquiry often relies on tests of significance or hypothesis-testing techniques when describing differences in data. Significance tests ultimately make a claim about differences, assuming at first that there are none and then tests this null hypothesis. The resulting observations may then be surprising (significant) or not, given the assumption that there are no differences. Significance tests such as chi-square, log likelihood and the Fisher Exact Test are often used in linguistic comparisons and are a type of keyness statistic of which the results are expressed in the form of an interpretable *p*-value. WordSmith Tools, for example, contains a keyness function where the output is based either on a chi-square or log likelihood statistic; the user may select which. Online calculators (such as the one used in this investigation) are often test-specific, where a preferred keyness metric, such as a log likelihood or a chi-square calculator, can be run independently from one another online.

The use of hypothesis-testing measures, however, has become increasingly problematic in research. The *p*-value and what it represents is often misunderstood, leading to incomplete data descriptions and incorrect conclusions (Brezina 2018; Cumming, 2014; Cumming, 2012; Halsey, 2019; Kilgarriff, 2005; Oakes, 1986). Its relevance and usefulness for linguistic comparison have also been brought into question (Kilgarriff, 2005). The criticism raised by Kilgarriff (2005), for instance, has in part to do with the assumption of no difference (by way of the null hypothesis). He states that there will always be some differences in language that are never truly random and that there are reasons certain language choices are exercised (Kilgarriff, 2005). Additionally, when large enough differences exist between frequencies, statistically significant results are eventually inevitable (Kilgarriff, 2005). In terms of answering the research question, it might be more useful to assume differences, regardless of how surprising they are, and to describe the nature of the differences in terms of magnitude. This is where effect size metrics play a role.

3.4.2 Measures of effect size

Linguistic research has called for the use of alternative keyness measures when describing differences (and similarities) in data, which are more transparent, meaningful and intuitive than significance tests (Brezina, 2018; Ellis, 2010; Gabrielatos, 2018; Gabrielatos & Marchi, 2012; Kilgarriff, 2009; Pojanapunya & Watson Todd, 2018). Generally, effect size is a statement regarding the presence or absence of an effect and the extent to which some type of intervention occurred. Ellis (2010:4) defines effect size as:

> the result of a treatment revealed in a comparison between groups (e.g., treated and untreated groups) or it can describe the degree of association between two related variables (e.g., treatment dosage and health).

Where significance tests do not make any claims about the nature, importance or the size of differences, effect size metrics are employed to help reveal something about the magnitude of the difference between groups (Ellis, 2010:5). Considering the structure of the data (single frequency count per pattern per register), where there are no means and standard deviations involved (Ellis, 2010:10), classic effect size measures such as Cohen's *d* and Pearson's *r* correlation would be inappropriate (Gabrielatos, 2018:235). Additionally, there is no effect in the literal sense of the above definition applicable here; that is, one value does not affect another. Rather, the concept of effect size as a measure of magnitude (as has recently been incorporated into linguistic corpus comparison studies) is used to describe the relationship between normalised frequencies of linguistic features (Gabrielatos, 2018; Gabrielatos & Marchi 2012; Kilgarriff, 2009).

A few effect size measures have been developed (some of which are implemented on the Excel spreadsheet via the UCREL site) to help describe the relationship between frequencies. These can also be ranked according to size. Sorting the differences based on size is useful when trying to establish the range of the differences, since the range of all the items needs to be considered before judging the size of the difference. The size of the difference is interpretable as large or small only in terms of the size differences of the rest of the output.

The metric that Gabrielatos (2018) and Gabrielatos and Marchi (2012) advocate is known as a measure of the percentage difference and is a straightforward measure of the distance between two normalised frequencies. This measure therefore provides a method of investigating not only differences but also similarities which, according to Taylor (2018:20), are often neglected in linguistic inquiry. The percentage difference between two normalised frequencies is calculated using the following formula (Gabrielatos, 2018:236):

$$\%DIFF = \frac{(NFC1-NFC2) *100}{NFC2}$$

Here, NFC1 is the normalised frequency count for the study corpus (corpus 1) and NFC2 is the normalised score for the reference corpus (Gabrielatos, 2018:236). It is immediately apparent (through the division by NFC2) that the percentage difference value will vary depending on which corpus is viewed as the target corpus and which the reference corpus or 'Corpus 1' and 'Corpus 2'. However, the size or magnitude of the difference is the same regardless. In terms of interpreting the values, Gabrielatos (2018:236) states that a percentage difference value of 100 is equal to twice the frequency of the key item in one corpus compared to the other. With every increase of 100, one is added to the difference, that is, if a value has a score of 200, that register has three times the number of that item than the register in Position 2.

The measure is directional and therefore the percentage difference calculation produces both positive and negative values based on whether lower frequencies occurred in Subregister 1 or Subregister 2 (Gabrielatos, 2018:236). Lower frequency in Position 1 will produce negative values, while lower frequency in Position 2 will produce positive values. Each set of 45 pairwise comparisons produced both mixed positive and negative values. These however cannot be sorted according to size, since the negative values are not on the same scale as the positive values i.e. negative values will sort below the positive values, even if they are larger than the positive value.

Since it is of no consequence here which is the study corpus and which the reference corpus (study Subregister and reference Subregister), and for the purposes of sorting values on the same scale, two regular expressions were devised before sorting the values. The first expression simply inverted the position of the subregisters (above and below the equation line) resulting in a positive score if it had been negative initially. The second printed the positive value out of the two individual columns, and could then be sorted according to size. These rankings can be interpreted on the same scale, where the register with the highest relative frequency between the two registers in each pair, is treated as Corpus 1.

Interpreting effect size differences as large or small depends on the range of all the values within the set of pairs in the data sheet (Gabrielatos & Marchi, 2012). For instance, the percentage difference score of 100 would be considered small if all other values were close to this same value. If a larger number of values deviated from 100, for instance, ranging from a percentage difference of 10,00 to 100,00, a percentage difference of 100,00 would be considered large. A percentage difference value is relatively large given the range of scores and each density type (premodifiers, phrasal postmodifier, clausal postmodifier, complement clause) are treated separately, wherein the range of values inform the interpretation of the differences observed

between subregisters. This also applies to each individual feature, where they are viewed in isolation.

3.5 Conclusion

Differences in online and offline registers is now measurable in terms of relative frequencies and effect size values. This chapter has set out the data and method employed in undertaking the data analyses. The method largely draws on keyness measures to compare and interpret the differences between registers with regard to noun phrase density devices. Usually, key *words* are identified through the use of statistical measures such as significance testing or effect size metrics, but in this study, the key items (that is, noun modifiers and complementisers) have already been identified. Therefore, instead of referring to key *words*, the application is extended to key *constructions*. The aim is to apply keyness techniques to these constructions and establish to what extent they identify or characterise each register. Given the fact that individual texts are not available through the online interface, the data really does not allow other types of descriptive statistics other than frequency and the effect size measures used here. The register classification as undertaken by Biber and Egbert (2016) is accepted wholesale. Estimated frequencies and noun counts form the basis of the descriptions and are compared across registers in terms of spoken and written languages, but also in terms of online and offline comparisons. The effect size metric *percentage difference* is used to describe the pairwise comparisons in terms of differences and similarities between the selected subregisters.

CHAPTER 4 ANALYSES

4.1 Introduction

This chapter quantitatively compares and describes the normalised frequencies of noun modifier and complementiser types across nonweb-based and web-based registers. The estimated frequencies for each individual grammatical pattern were normalised to a base of 100 000 nouns. The normalised frequencies presented in the graphs are mostly brought into context with some reference to noun phrase feature patterns in the *LGSWE* and findings from studies focused on noun phrase structure. Overall, three main comparisons are made regarding densification devices: first, between the *spoken* component of COCA and the *written* registers in both COCA and CORE; second, between online and offline registers; and third, between two given registers by way of pairwise comparisons. These comparisons are made for each densification device following the cline of densification (that is, prenominal modifiers, postnominal phrases as modifiers, postnominal clauses as modifiers and noun complement clauses).

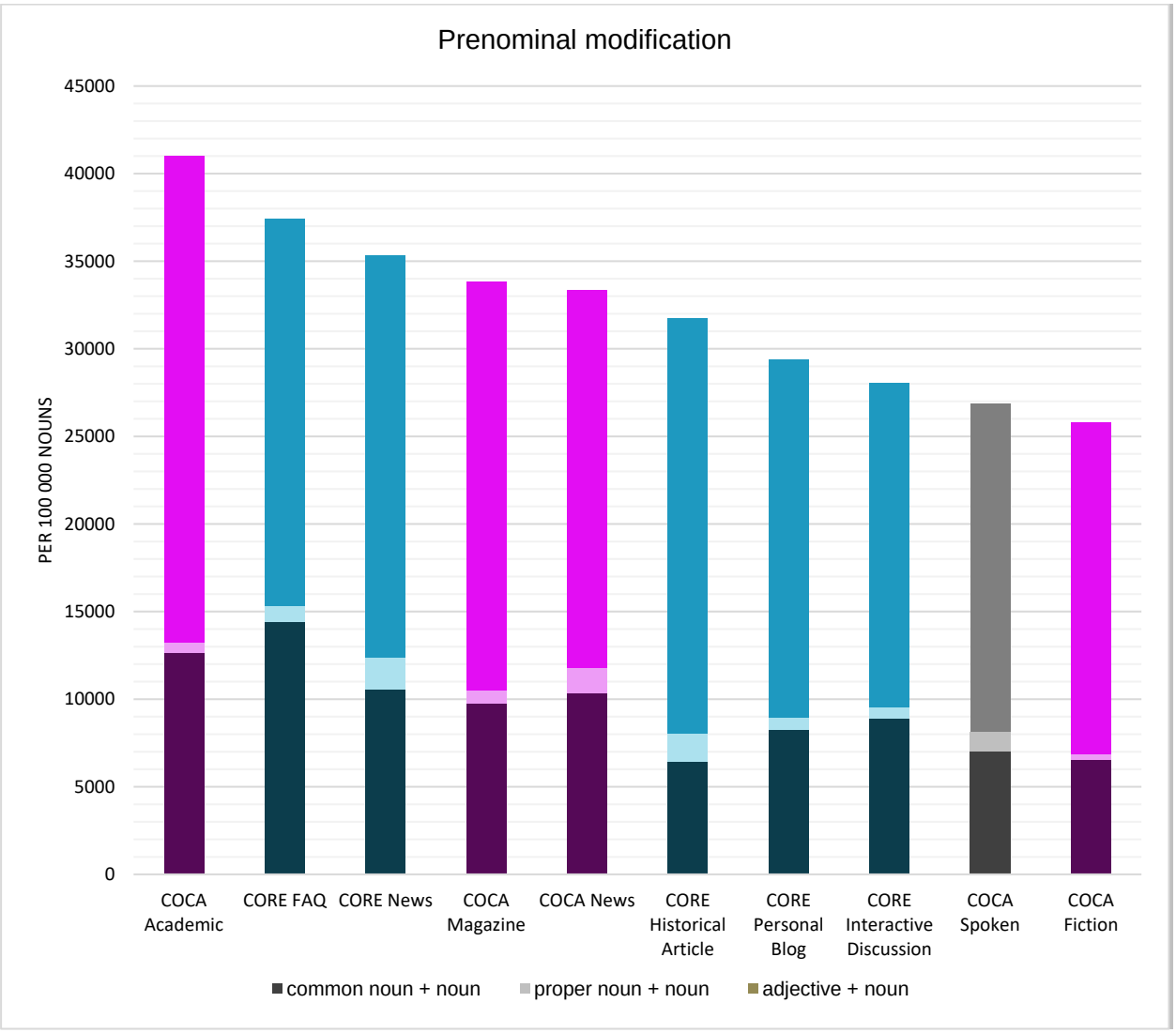
4.2 Prenominal modifiers

Prenominal phrasal modifiers are an important feature of noun phrase density. Studies have shown that higher frequencies of nominal premodifiers distinguish written fiction and nonfiction registers from spoken discourse (Biber *et al*., 1999:589; Biber & Gray, 2011:299–230; Gray, 2015:63). Further, MD analyses also indicate that dimensions marked with an information purpose have high factor scores for nominal premodification features (Biber, 1988). Premodifiers are especially frequent in news and academic prose, less frequent in fiction and very infrequent in conversation (Biber *et al*., 1999:589). In addition, important distinctions are made in terms of individual premodifier type. Among all types of nominal premodifiers, adjectives are the most common type, being very frequent in both fiction and nonfiction writing and the most common in academic writing (Biber *et al*., 1999:589). Common nouns are the second-most common modifier type. They are important to the dense packaging of referential information within a single phrase expressing the logical relationship between the lexical constituents within that phrase (Biber *et al*., 1999:589–590). Distribution patterns in the *LSGWE* show that news discourse and expository writing are heavily marked by premodifying nouns, while they are less common in fiction and conversation.

Figure 4-1 below presents the relative frequencies of nominal premodification types for offline and online registers per 100 000 nouns. For the purposes of ranking the registers from highest to lowest frequency, the frequencies of the three individual premodifier types are combined. Therefore, individual patterns do not create the cline from higher to lower scores. Colour is used

to distinguish between online (teal) and offline (purple) written registers, while Spoken is grey. The three types of premodifier are distinguished by tone bands, with common nouns as the darkest tone, proper nouns as the lightest tone and adjectives as mid-tone.

Figure 4-1: Distribution of prenominal modifiers per 100 000 nouns



First, this figure shows that except for Fiction, features of noun phrase density are overall more common in written registers than in Spoken, making written text largely distinctive from this conversational component. In particular, the frequencies of proper noun modifiers in Spoken are more pronounced compared to most written registers; its premodifying adjectives and common nouns occur at lower frequencies.

It is evident from this graph that prenominal modification is relatively frequent in all of the registers. It is the most frequent in Academic, followed by online FAQ and online News. The lowest relative frequencies are in Fiction and Spoken. The overall ranking indicates that the registers with the

highest proportion of premodifiers are informational writing, while registers that are more narrative in nature tend to have fewer premodifiers. The following text excerpt from FAQ illustrates the dense use of premodifiers (underlined) in service of referent specification. Head nouns (including those with no premodifier) are in bold:

Text 1:

> *asbestos is a naturally occurring mineral* ***fibre*** *that was widely used in the building* ***industry****. A residential* ***building*** *constructed or refurbished between 1950 and 1985 or a commercial* ***building*** *built before 2004 is likely to contain asbestos containing* ***material****.*
>
> ***Asbestos*** *may be found on construction* ***sites*** *in* ***products*** *such as:*
>
> *cement* ***sheeting***
>
> *corrugated* ***sheeting***
>
> *flat sheeting* ***panels***
>
> ***shingles*** *or* ***tiles*** *(external or ceiling)*
>
> ***pipes****,* ***tubes*** *or* ***fittings***
>
> *compressed* ***asbestos sheeting*** *such as* ***millboard***
>
> *floor* ***coverings*** *such as vinyl* ***asbestos tiles***
>
> *textured* ***paints*** *and* ***coatings***
>
> ***compounds***
>
> ***mastics****,* ***sealants****,* ***putties*** *and* ***adhesives***
>
> *electrical* ***panel partitioning***. (FAQ, CORE)

Registers such as Spoken and Fiction, which have lower proportions of modifiers, rely less on referent specification, since the need to elaborate on these structures to the extent that, for instance, Academic or FAQ do is largely unnecessary where there is an immediate shared context, and the purpose of any popular narrative is to retain the reader's attention. The following descriptive or narrative excerpt is taken from Fiction and shows how nouns (in bold) are less likely to have premodifiers (underlined) when they occur:

Text 2:

> *We did not go for a* ***walk*** *on the first* ***day*** *of the* ***year****. The Christmas* ***snow*** *had melted, and* ***rain*** *had been falling since* ***dawn****, darkening the* ***shrubbery*** *and muddying the* ***grass****, but that would not have stopped my* ***aunt*** *from dispatching us. She believed in the* ***benefits*** *of*

fresh ***air*** *for* ***children*** *in all* ***weather****. Later, I understood, she also enjoyed the* ***peace*** *and* ***quiet*** *of our* ***absence****. No, the* ***cause*** *of our not walking was my cousin Will, who claimed his* ***cold*** *was too severe to leave the* ***sitting room sofa****, but not so bad that he couldn't play* ***cards****. His* ***sister*** *Louise, he insisted, must stay behind for a* ***game*** *of racing* ***demon****. I overheard these* ***negotiations*** *from the* ***corridor*** *where I loitered, holding my* ***aunt's*** *black* ***shoes****, freshly polished, one in each* ***hand****.* (Fiction, COCA)

While online FAQ and News have higher proportions of premodifiers, more offline registers contribute to the higher end of the frequencies seen on this graph. The majority of online registers, and specifically Historical article, Personal blog and Interactive discussion, rank in the lower-frequency range.

Off all premodifier types, adjectives are the most common density device in all registers compared to premodifying common nouns and proper nouns. Further, they are the most frequent in Academic, followed by Historical article and the least frequent in Interactive discussion.

Distribution patterns for adjectives as premodifiers show that the Spoken register in COCA is not overtly distinguished from other written registers in the data. Spoken has a moderately higher frequency than online written Interactive discussion, and Spoken is also very similar to offline Fiction. Spoken is, however, more distinguished from Academic, Historical article and Magazine, which overall have higher proportions of adjectival modification.

In terms of offline and online written registers, premodifying adjectives place online registers squarely in a higher relative-frequency range. Fewer online registers appear in the lower-frequency positions compared to the offline COCA texts with overall lower frequencies.

Premodifying common nouns are markedly less frequent than premodifying adjectives across all registers. Additionally, Spoken in COCA is not distinguished from written registers in terms of this device.

Further, FAQ, followed by Academic, has the highest relative frequency of premodifying common nouns, while online Historical article has the lowest occurrences. Even though online FAQ and online News (the third highest frequency of combined premodification devices) have higher proportions of common noun premodifiers; more offline COCA registers occur in the higher-frequency range and more online registers appear in the lower range.

Proper nouns are the least frequent modifying type of all three nominal premodification devices. They are more common in online News and Historical article and are the least frequent in Fiction.

While Spoken has shown little distinction from written registers for both modifier types described so far, this register is even less distinct from writing in terms of premodifying proper nouns. When ranking registers according to proper noun as premodifier from highest to lowest relative frequency, there are proportionately more premodifying proper nouns in online texts than in offline registers. Lower frequencies are evident in Personal blog, Interactive discussion and Academic. Spoken, both online and offline News and Historical article have higher occurrences of premodifying proper nouns.

As a short summary of normalised frequencies of premodification types, the data in the graph show that premodification devices are common in online and offline written registers. While Spoken has lower relative frequencies of premodifier types, there is little distinction between this register and the other written components. This observation is the most pronounced in premodifying proper noun patterns specifically.

Premodifying adjectives occur more commonly in online registers than in offline writing. In terms of written online and offline registers, higher relative frequencies of adjectives mark Academic, Historical article, Magazine and CORE News. Common nouns are more frequent in offline registers, but especially pronounced in information-driven writing, including FAQ, Academic and CORE and COCA News when compared to Interactive discussion, Personal blog, Fiction and Historical article.

Further, common nouns as premodifiers make an important general distinction between fiction and nonfiction; adjectives are most common in online nonfiction and less so in offline nonfiction and least common in online involved writing. As a feature of noun phrase density, phrasal premodification is an important strategy in distinguishing online written registers from offline written registers in terms of two particular modifier types, with density features occurring more frequently in online registers.

The following subsections provide the percentage-difference scores for each of the three individual premodifier types below and describe the most different – and, in some cases, the most similar – register types in terms of magnitude.

4.2.1 Proper nouns

The previous graph (Figure 4-1) containing normed frequencies showed that adjectives in particular are especially important to density in online writing and that proper nouns are the least common. Even though frequencies of premodifying proper nouns are evidently fewer than that of adjective and common noun premodifiers, proper nouns mark the largest differences between the registers in quantitative terms. In Table 4-1, the register pairs are arranged from most different to

most similar, based on the effect size measure of percentage difference. The corresponding ratio of relative frequency is included to translate the percentage-difference score in terms of ratio. The online registers reflect the colour codes applied in Figure 4-1, with online CORE registers in teal, offline written COCA registers in purple and Spoken in grey. The register to the left of the pair contains the higher relative frequency and the register on the right the lower frequency.

Table 4-1: Most different register pairs for proper nouns as prenominal modifier

Register pairs	%DIFF score	RRF
CORE News_Fiction	447,77	5,48
Historical Article_Fiction	375,67	4,76
News_Fiction	324,85	4,25
Spoken_Fiction	236,50	3,37
CORE News_Academic	199,30	2,99

Table 4-1 shows the registers that are most distinctive in their use of premodifying proper nouns. The normalised frequencies from Figure 4-1 shows that this device is the least common among all premodifier types, but it has the largest percentage difference range of all three prenominal modifiers, with premodifying proper nouns occurring five and a half times as often in CORE News than in Fiction, and over four and half times as often in Historical article than in Fiction.

For this particular feature, Spoken differs most from Fiction, with occurrences of over three times more common in Spoken than in Fiction. While Spoken differs from Fiction, Spoken is similar to written registers (these scores are not included in this table), including FAQ (20,43), offline News (26,25), Historical article (41,36) and Magazine (45,83). Slightly larger differences involve online News (62,79) and Personal blog (64,58), and the difference between Spoken and Academic (83,86) is only slightly larger than the difference between Spoken and Interactive discussion (83,49). These scores reflect the extent to which proper nouns differentiate Spoken from the other written registers. This is not to say Spoken is similar to all written registers, but on average the distinctions are not especially prominent when these scores are compared to the difference between Spoken and Fiction.

The larger differences in Table 4-1 show that written online registers are distinct from offline Fiction. In the proper noun patterns in the data (not visible in this table) the 18 register pairs that differed most have values exceeding a percentage difference of 100,00 (being twice as frequent). Within this group, 11 register pairs show contrasts between written online and offline registers.

4.2.2 Common nouns

Premodifying common nouns are the second most common premodifier type in Figure 4-1. The higher frequencies in FAQ and Academic contrast with the lower relative frequencies in Historical article, Fiction and Spoken. These registers feature in the most different register pairs for prenominal common nouns, as set out in Table 4-2 below. The registers marked in teal belong to CORE, the registers in COCA are marked in purple and Spoken is presented in grey. The register to the left of each pair has the higher frequencies of the two. The percentage difference score is listed in the second column, and the ratio in the last column is a measure of the ratio difference between each register pair for premodifying common nouns. (See Table A.1 in the Appendix.)

Table 4-2: Most different register pairs for common nouns as prenominal modifier

Register pairs	%DIFF score	RRF
FAQ_Historical Article	123,66	2,24
FAQ_Fiction	121,05	2,21
FAQ_Spoken	104,99	2,05
Academic_Historical Article	96,07	1,96
Academic_Fiction	93,78	1,94

The range of the differences between registers is smaller than that of proper noun premodifiers. The highest difference is marked between FAQ and Historical article, where common nouns are modified more than twice as much in FAQ. Further, FAQ marks similar size differences with Fiction and with Spoken respectively. Slightly smaller differences exist between Academic and Historical article and between Academic and Fiction, where premodifying common nouns are more common in Academic in both pairs. The most similar among all the register pairs (not explicitly shown in Table 4-1) is Fiction and Historical article (1,18). These two registers are practically equal in their usage.

Further, Spoken is distinctive from FAQ, as shown in Table 4-2, but it is also different from Academic (79,70), online News (CORE) (50,23) and offline News (47,36). The differences between Spoken and other written registers become smaller along this cline. Premodifying common nouns occur less than one and a half times as frequently in Spoken than in Magazine (38,44), Interactive discussion (26,69) and Personal blog (17,37). They are also slightly more frequent in Spoken than in Historical article (9,11) and Fiction (7,83).

In terms of offline and online register comparisons, eight of the top 15 register pairs with a percentage difference score of 50,00% or more are contrasts between offline and online texts; the remaining differences are intercorpus-related. In this particular set of contrasting pairs, half the pairs indicated a higher frequency in online writing and half in offline writing. When these online–offline differences are ranked according to effect-size differences from largest to smallest, 75% of the pairs (which are in the top half) show that the higher frequencies occur in online registers.

Table 4-1 further shows that larger differences exist between informational writing (FAQ and Academic) and narrative texts (Fiction and Historical article). Larger differences (not shown in Table 4-1) are also found between online News and Historical article (63,91) and online News and Fiction (62,00). Additionally, online News (CORE) and offline News use premodifying common nouns in practically the same way (1,94), where offline News has the slightly higher frequency rate of the pair.

Another noteworthy comparison is the distance between Academic and both News registers. With just over one times as many occurrences in Academic than in News, at 19,62 the difference to online News (CORE) is slightly smaller than the relation to offline News (21,95). Both these scores are relatively small compared to the effect size differences of the most different register pairs.

4.2.3 Adjectives

Figure 4-1 shows that premodifying adjectives are the most frequent occurring premodifier type across all registers. While a gradual formal-to-informal cline is evident in the ranking of registers (Figure 4-1), the differences in frequencies between registers do not exceed 50%. This is indicated in Table 4-3 below, which lists the most different register pairs of adjectives as prenominal modifier. Following this is Table 4-4, which provides the effect size values of the most similar registers ranked from the smallest to largest differences.

The same formatting and layout aspects apply in both tables. The first column contains the relevant register pairs, where the register to the left has higher frequencies than the register to the right. The second column indicates the percentage difference score for each of the individual register pairs ranked from larger to smaller differences. The registers are colour coded, where teal represents online registers in CORE, purple indicates offline COCA registers and grey indicates Spoken. (See Table A.2 in the Appendix.)

Table 4-3: Most different register pairs of adjectives as prenominal modifier

Register pair	%DIFF score	RRF
Academic_Interactive Discussion	50,02	1,50
Academic_Spoken	48,53	1,49
Academic_Fiction	46,66	1,47
Academic_Personal Blog	36,04	1,36
Academic_News	28,78	1,29

As is evident in Table 4-3, the largest contrast lies between offline Academic and online Interactive discussion. Academic largely displays the most obvious contrasts as is evident from the table; the differences between online and offline registers become smaller along this cline. Only 10 register pairs differ with a percentage difference equal to or greater than 25,00, and of these, six pairs are offline–online-related, while the remaining four pairs are interregister contrasts.

The second-largest difference involves Spoken and Academic. However, differences between Spoken and other written registers, including Historical article (26,75), Magazine (24,74), CORE News (22,67), FAQ (18,10) and offline News (15,34) are less dramatic. It would seem that Spoken is more similar to these written registers.

Table 4-4: Most similar register pairs of adjectives as prenominal modifier

Register pair	%DIFF
Spoken_Interactive Discussion	1,00
Fiction_Spoken	1,27
Magazine_Historical Article	1,61
Magazine_CORE News	1,68
Fiction_Interactive Discussion	2,29

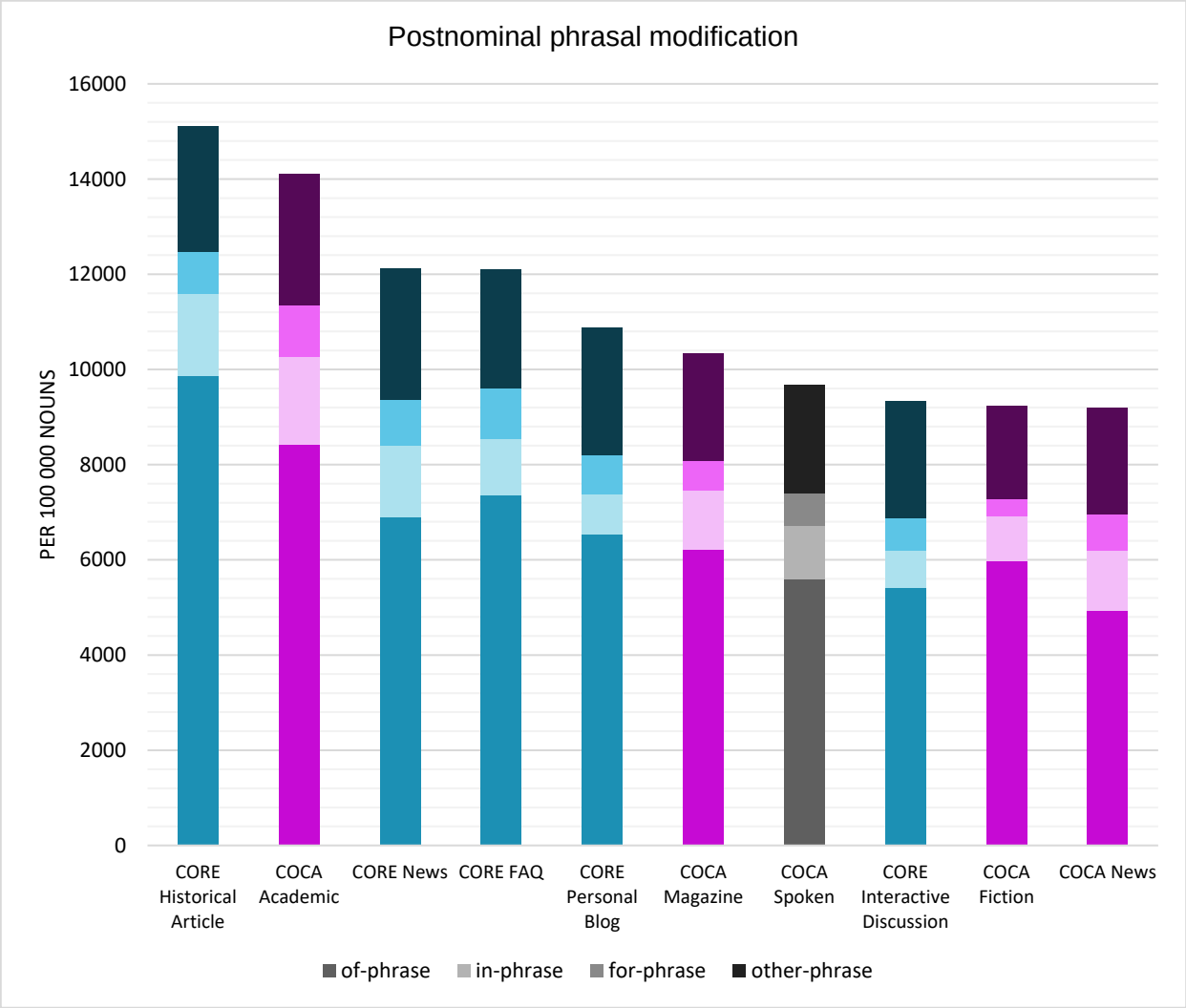
As can be seen in Table 4-4, Spoken and online Interactive discussion are most similar, followed by Fiction and Spoken. All the pairs shown here are practically the same in their distribution of premodifying adjectives. In fact, the bottom 17 register pairs in this pattern have a difference score of less than 10%, and another 18 pairs' scores range only from 10,28% to 24,74% (this leaves the top 10, including the five pairs shown in Table 4-3, with a range from 25,16% to 50,02%). Therefore, while larger differences obviously exist between the most opposing pairs, there are

fewer of these types of frequency differences for this pattern across all registers compared to the contrasts of other two premodifier types. (See Table A.3 in the Appendix.)

4.3 Postnominal phrasal modifiers

This section discusses data patterns related to postnominal phrasal modification in online and offline registers. Patterns are described in terms of differences between Spoken and other written registers as well as online and offline contrasts. In addition to nominal premodification, phrasal postmodifiers are a distinguishing feature of noun phrase density. Figure 4-2 provides the relative frequencies (per 100 000 nouns) of the three most common preposition phrases as nominal postmodifier in this particular data set. The prepositions *of*, *in* and *for* (in this order) are the three most frequently occurring structures introducing phrasal postmodifiers in noun phrases. The remaining nine prepositions are grouped in the column marked ‘other-phrases’ (*about, at, between, by, from, like, on, to* and *with*), which appears at the top of each stacked column. Colour is used to distinguish online registers from offline registers (online written is teal, offline written is purple and offline spoken is grey), while tone is used to distinguish between the prepositions (*of*-phrases are mid-dark, *for*-phrases are mid-light, *in*-phrases are light and other prepositional phrases are dark).

Figure 4-2: Postnominal phrasal modification per 100 000 nouns



First, considering the ranking scale of registers shown here, it is evident that the written registers, in general, have higher frequencies of phrasal postmodification compared to Spoken. Online Historical article has the highest proportions of combined phrasal postmodification, followed by Academic. Offline News has the lowest relative frequency.

Further, apart from more extreme contrasts with Historical article and Academic, for instance, Spoken is not obviously distinguished from offline or online written texts in terms of any individual phrasal postmodifier type or combined frequencies of all phrasal postmodification devices.

Figure 4-2 also shows lower overall frequencies of phrasal postmodification patterns compared to premodification patterns (Figure 4-1), but these density devices are still fairly common in both the CORE and COCA registers here. Second, *of*-phrases are singularly the most frequent phrasal postmodifier in both the written registers and spoken component. The distribution patterns of *of*-phrases in the *LGSWE* show that *of* accounts for more than 60% of phrasal postmodifier types,

and is more common in expository written registers than in popular registers such as fiction, drama and conversation (Biber *et al*., 1999:635; Biber & Clark, 2002:58–61).

Figure 4-2 also shows that phrasal postmodification as a densification strategy is most apparent in online written registers, where Historical article, online News, FAQ and Personal blog are positioned on the higher-frequency end of the ranked ordering scale. More offline registers contain lower relative frequencies of phrasal postmodification structures; Academic is the only offline register with relatively high frequencies compared to other offline and online registers.

Historical article has the highest combined frequencies of phrasal postmodification (driven in large part by *of*-phrases) and surpasses overall relative frequencies in offline Academic. Even though Academic has slightly higher frequencies of *in*-phrases, *for*-phrases and *other*-phrases, it contains fewer *of*-phrases. While Academic is classically expository in style and purpose, Historical article is strongly marked for prepositional phrases on dimensions characterised as literate–informational and informational narration respectively (Biber & Egbert, 2016:108). This information might place this register into perspective despite it being generally categorised as narrative. The importance of *of*-phrases as a densification device in Historical article in particular is illustrated by the following excerpt, where most noun heads contain both premodifiers and postmodifiers, but few nouns have no postmodifier. Noun heads are in bold; phrasal postmodifiers are underlined:

Text 3:

> *The Cavendish Laboratory has an extraordinary* ***history*** *of discovery and* ***innovation*** *in Physics since its* ***opening*** *in 1874 under the* ***direction*** *of James Clerk Maxwell, the University's first Cavendish* ***Professor*** *of Experimental Physics. Up till that* ***time****,* ***physics*** *meant theoretical* ***physics*** *and was regarded as the* ***province*** *of the mathematicians. The outstanding experimental* ***contributions*** *of Isaac Newton, Thomas Young and George Gabriel Stokes were all carried out in their* ***colleges****. The* ***need*** *for the practical* ***training*** *of scientists and engineers was emphasised by the* ***success*** *of the Great* ***Exhibition*** *of 1851 and the* ***requirements*** *of an industrial society. The* ***foundation*** *of the Natural Sciences Tripos in 1851 set the* ***scene*** *for the* ***need*** *to build dedicated experimental physics* ***laboratories*** *and this was achieved through the* ***generosity*** *of the Chancellor of the University, William Cavendish, the Seventh* ***Duke*** *of Devonshire. He provided £6,300 to meet the* ***costs*** *of building a physics* ***laboratory****, on* ***condition*** *that the Colleges provided the* ***funding*** *for a* ***Professorship*** *of Experimental Physics. This led to the* ***appointment*** *of Maxwell as the first Cavendish* ***professor****.*

*The **move** was completed in 1974 and a completely new **phase** of discovery began. Large **facilities** were developed in radio **astronomy** and semiconductor **physics**, which continue to be frontier **areas** of research within the Laboratory. Completely new **disciplines** were fostered. With Sam Edward's **appointment** as Pippard's successor in 1984, soft condensed **matter** became a major **component** of the Laboratory's programme. This led in **turn** to major **initiatives** in biological physics and the **physics** of medicine. Polymer semiconductor **physics** has flourished under Edwards' **successor** Richard Friend. In the first **decade** of the 21st century, new **frontiers** have been opened up in the **areas** of nanotechnology, cold atoms and ultra-low temperature physics* (Historical article, CORE).

Some further noticeable observations pertain to offline News (COCA) and online News (CORE). The first is the obvious lower frequencies in offline News compared to online News in both premodification and phrasal postmodification patterns. Phrasal modification therefore seems to consistently characterise online News in terms of density, more so than it does offline News. Second, as shown in Figure 4-2, offline News has the lowest overall frequencies of phrasal postmodifiers across all registers, which is mainly due to lower uses in *of*-phrases. (*In*-phrases, *for*-phrases and *other*-phrases are not the least frequent in this register compared to most of the other registers in the data set.) The following text is a typical example of offline News. In the text, all noun heads are in bold, but head nouns that contain no phrasal postmodifier are in bold and underlined; very few nouns contain a phrasal postmodifier and these structures are seemingly avoided:

Text 4:

*When Paula Symons joined the U.S. **workforce** in 1972, **typewriters** in her office clacked nonstop, **people** answered the **telephones** and the hot new technology revolutionizing **communication** was the fax **machine**. Symons, fresh out of **college**, entered this brave new **world** thinking she'd do pretty much what her parents' **generation** did: Work for just one or two **companies** over about 45 years before bidding farewell to **co-workers** at a retirement **party** and heading off into her sunset **years** with a **pension**. Forty years into that **run**, the 60-year-old communications **specialist** for a Wisconsin-based insurance **company** has worked more than a half-dozen **jobs**. She's been laid off, downsized and seen the **pension** disappear with only a few thousand **dollars** accrued when it was frozen. So, five years from the age when people once retired, she laughs when she describes her future **plans**. 'I'll probably just work until I drop,' she says, a **sentiment** expressed, with varying **degrees** of humor, by numerous **members** of @@@@ Baby Boomers, Symons and her **husband** had the **misfortune** of approaching retirement **age** at a **time** when stock **market** crashes diminished their 401(k) nest **eggs**, **companies** began eliminating defined*

*benefit **pensions** in record **numbers** and previously unimagined technical **advances** all but eliminated entire job **descriptions** from travel agent to telephone operator. At the same **time**, **companies** began moving other **jobs** **overseas**, to be filled by **people** willing to work for far less and still able to connect to the U.S. **market** in real **time*** (News, COCA)

Offline News seems to avoid the use of phrasal postmodification devices where it has the option, grammatically speaking, to incorporate them into the text. Restrictions in terms of word count and writing space are certainly considerations that influence the offline print form, possibly leading to lower frequencies in phrasal postmodification devices in this register.

Evidently, *in*-phrases and *for*-phrases are less common than *of*-phrases and, considering that *other*-phrases are made up of the remaining nine prepositions combined, the frequencies for these individual structures are rare in comparison (Figure 4-2). Nominal postmodifying *in*-phrases and *for*-phrases are especially common in scientific prose compared to other general nonfiction, fiction and involved registers (Biber *et al*., 1999:635, Biber & Clark, 2002:62; Biber & Gray, 2011:241; Biber & Gray, 2012:323–324).

As shown in Figure 4-2, these phrasal postmodifying structures are more common in written registers than in Spoken, but as in the case of *of*-phrase postmodification, Spoken neither has the lowest relative frequency nor a rate which is that much different from general written registers. Spoken has a relatively high frequency of *for*-phrase and *in*-phrase postmodifiers, compared to written Fiction, Interactive discussion, Personal blog and Magazine. Rather, the rankings on the scale show a clearer distinction between Spoken and informational registers such as Academic and Historical article, which are more pronounced. Therefore, while phrasal postmodification definitely marks most written online and offline registers, Spoken is not obviously different from popular and involved registers.

Figure 4-2 further shows that offline Academic has the highest relative frequency of both *in*-phrases and *for*-phrases. Of the online registers, Historical article, News (CORE) and FAQ have higher frequencies of these postmodifiers than Personal blog and Interactive discussion. Overall, *in*-phrase postmodifiers are more common in offline registers than online registers, but the two online registers that appear in this upper frequency range, Historical article and online News, have higher frequencies than Magazine and offline News. And, even though offline News (COCA) has slightly higher frequencies of *in*-phrase postmodifiers here, other phrasal postmodifier types fall in the lower-frequency range for this register.

Further, *for*-phrases are markedly more common in online registers and have higher relative frequencies than most of the offline texts. These postmodifiers seem to create a relatively clear

cline from expository writing to more general nonfiction to fiction. *For*-phrase postmodifiers in FAQ (CORE), for instance, follow nouns such as *need*, *details*, *application*, *contract*, *standards*, *requirements*, *guidelines*, *infrastructure*, *tool*, *opportunity*, *responsibility*, *cause*, *liability*, *choice*, *deadline*, *petition*, *evidence*, *supply*, *referral*, *investment* and *request.* These nouns would often appear in titles and names of places and organisations and technical topics such as:

1. Registrar's **Certificate** for Burial or Cremation
2. The medical **term** for high blood pressure is hypertension
3. An employee who believes he or she has **grounds** for a personal grievance against you
4. The list of products under the **scope** for the Use of Foreign Review pilot phase of the project

These head nouns (many of them nominalisations) can broadly be characterised as abstract rather than concrete and are in a sense detached, as opposed to those of Fiction. Nouns often occurring in the concordance sample data of Fiction include animals (*worms*, *horses*) and people (*columnist*, *reporter*, *manager*, *engineer* and *volunteer*) as well as *time*, *thing*, *payback*, *prayer*, *reputation*, *loss*, *affection*, *fondness*, *need*, *obsession*, *gift*, *taste*, *cure*, *passion*, *greed*, *love*, *search*, *mercy*, *veneration* and *hopes.* Phrasal postmodification serves Fiction (COCA) in a different way than it does FAQ, where nouns, some also abstract but including more concrete entities, are more personalised and appeal to emotion rather than being indifferent:

5. He was a **sucker** for a pretty face and sweet smile
6. They've always had a **reputation** for ruthlessness
7. She even has a great deal of **affection** for Pam
8. He had a **passion** for truth

The section above described the distribution patterns of phrasal postmodification in online and offline registers (Figure 4-2). As one of the most important distinguishing features of noun phrase density, phrasal postmodification is in large part more common in informational registers compared to general fiction and nonfiction; for the most part, it is online writing (informational and general registers) that has overall higher frequencies than most offline registers.

The following tables provide the percentage difference scores of the most different register pairs for postnominal *of*-phrases (Table 4-5), the most similar pairs of postnominal *of*-phrases (Table

4-6) and the most different pairs for *in*-phrases (Table 4-7) and *for*-phrases (Table 4-8), respectively. In each table, the register in the left position of the pair contains the highest relative frequency. The register pairs appear in the first column and the percentage difference scores in the second column. The last column contains the ratio of relative frequency that interprets the percentage difference score in ratio terms. Offline registers are marked in purple, online registers are in teal and Spoken is marked in grey to show contrasts between Spoken and the written registers and between online and offline registers. (See Table A.4 in the Appendix.)

Table 4-5: Most different register pairs for postnominal *of*-phrases

Register pairs	%DIFF	RRF
Historical Article_News	100,48	2,00
Historical Article_Interactive Discussion	82,28	1,82
Historical Article_Spoken	76,19	1,76
Academic_News	70,87	1,71
Historical Article_Fiction	65,37	1,65
Historical Article_Magazine	58,66	1,59
Academic_Interactive Discussion	55,36	1,55
Historical Article_Personal Blog	51,02	1,51
Academic_Spoken	50,17	1,50
FAQ_News	49,45	1,49

The *of*-phrase data previously presented in Figure 4-2 show a proclivity for online registers (informational and general) to diverge from offline registers in terms of their ranked positions. Table 4-5 presents the highest contrasts between registers, where 10% of register contrasts show a difference larger than 50% between register pairs. Evidently, Historical article is repeatedly involved in the larger differences and it more commonly contrasts with offline registers. It also creates the largest percentage difference in *of*-phrase use with offline News, where *of*-phrase postmodifiers are more than twice as frequent in the online register than offline News.

Further, where the largest effect sizes are concerned, Table 4-5 indicates that offline–online contrasts are more common in *of*-phrase postmodification than interregister (offline–offline or online–online) contrasts. As in the case of premodification, Spoken is mostly distinct from the informational registers Historical article and Academic, where phrasal postmodifiers are more frequent in the written registers.

Table 4-5 also emphasises a divide between major information-driven registers and seemingly more general fiction and nonfiction, as well as involved written texts. *Of*-phrase postmodification is more than one and half times as frequent in informational writing (Academic, Historical article and FAQ) than in registers such as Personal blog, Interactive discussion and Fiction.

While *of*-phrases seem to create an offline–online distinction where the largest percentage difference values are concerned, Table 4-6 below presents the most similar register pairs of this same postmodification device. Online registers are marked in teal, offline registers in purple and Spoken in grey. The second column contains the percentage difference value of the size difference between each pair. In Column 1, the register to the left of each pair has the higher proportion of *of*-phrase postmodification of the two registers.

Table 4-6: Most similar register pairs for postnominal *of*-phrases

Register pairs	%DIFF
Spoken_Interactive Discussion	3,46
Magazine_Fiction	4,23
Personal Blog_Magazine	5,05
CORE News_Personal Blog	5,64
Fiction_Spoken	6,54
FAQ_CORE News	6,58
Personal Blog_Fiction	9,50
Interactive Discussion_News	9,98
Fiction_Interactive Discussion	10,22
CORE News_Magazine	10,98

There is virtually no difference in *of*-phrase postmodification use when comparing Spoken to Interactive discussion and Fiction (both written registers). Spoken is also similar to Magazine (11,05) and offline News (13,79), although these two effect sizes are not shown in this table. This table shows that while Spoken does not commonly appear in the most similar pairs, it is not different from fiction and nonfiction writing.

It further seems that density makes less of a distinction in offline–offline and online–online interregister pairs, but the offline–online differences are slightly more pronounced, though still similar. Phrasal density (at least for *of*-phrasal postmodification) seems equally important to online

involved and informational writing (online News and Personal blog) and to online News (CORE) and FAQ, which are arguably more information-driven in purpose. The more pronounced differences in Table 4-6 are between online and offline writing, where effect sizes become larger (albeit still small).

The other two most frequent phrasal postmodifiers, *in*-phrases and *for*-phrases – relatively infrequent compared to *of*-phrases – show larger effect-size differences between registers than *of*-phrases do. From Table 4-7 and Table 4-8 below it is clear that *for*-phrasal postmodification (the least frequent phrasal postmodifier) has the largest effect size differences between online and offline registers. (See Table A.5 in the Appendix.)

Table 4-7: Most different register pairs for postnominal *in*-phrases

Register pairs	%DIFF	RRF
Academic_Interactive Discussion	136,25	2,36
Historical Article_Interactive Discussion	121,45	2,21
Academic_Personal Blog	115,58	2,16
Historical Article_Personal Blog	102,08	2,02
CORE News_Interactive Discussion	92,22	1,92

In-phrases are evidently more than twice as frequent in written expository writing and general information-driven registers such as Academic and Historical article. Online News (CORE) has almost twice as many occurrences of *in*-phrase postmodifiers than Interactive discussion. These information-type registers contrast sharply with involved writing, including Interactive discussion and Personal blog. Academic has more than twice the occurrences of *in*-phrase modification than Interactive discussion and more than one and half times that of Spoken (66,40).

Despite the largest difference being between offline Academic and online Interactive discussion, it is mainly involved Interactive discussion that establishes contrasts with other online and offline registers, including Historical article (102,08), online News (92,22), but also FAQ (51,87) and offline News (63,21) (not listed in this table). It therefore seems that this densification device strongly distinguishes between text types, where involved writing is often contrasted with information-type texts.

While Academic shows a large contrast with involved-written Interactive discussion and Personal blog respectively, the data in Table 4-7 show that the larger differences occur more commonly between online, than offline–online contrasts do. Considering the data overall, 33% of register

pairs within this construction have an effect-size difference larger than 50%. More register pairs with an online–online or offline–offline component appear in the higher slots, whereas more offline–online pairs show the smaller differences. Where the largest differences appear between registers, density makes less of a distinction between offline and online writing as opposed to *of*-phrase postmodification and, as shown below, *for*-phrase postmodification. (See Table A.6 in Appendix for complete pairwise comparison.)

Table 4-8: Most different register pairs for postnominal *for*-phrases

Register pairs	%DIFF	RRF
Academic_Fiction	218,55	3,19
FAQ_Fiction	209,56	3,10
CORE News_Fiction	180,79	2,81
Historical Article_Fiction	150,73	2,51
Personal Blog_Fiction	133,36	2,33
News_Fiction	124,39	2,24
Spoken_Fiction	101,62	2,02
Interactive Discussion_Fiction	96,38	1,96
Magazine_Fiction	83,94	1,84

Table 4-8 shows that *for*-phrase postmodifiers (even though they are very rare compared to *of*-phrases) create some of the largest differences between registers. Evidently, the proportion of *for*-phrase postmodification is more than 80% larger in all other registers compared to Fiction, where the larger contrasts with online writing seem to follow a cline from informational to general and involved writing. It is interesting that the differences with online registers group in this way, when contrasts with offline registers show no particular pattern in terms of a formal–informal cline, except for the contrast with Academic. The expectation is that the difference between Spoken and Fiction would be smaller than, for instance, the difference between Magazine and Fiction and even Interactive discussion (being a written register) and Fiction. Not only does Spoken have twice the frequency of a density feature largely associated with the written form compared to Fiction, it is also very similar to other written registers such as Personal blog (15,74) offline News (11,30), Magazine (9,61) and Interactive discussion (2,67).

Overall, *for*-phrase postmodification distinguishes Fiction from online writing in a rather systematic way that is not reflected in contrasts with offline registers. Additionally, the differences between

register pairs containing an online component are, generally speaking, much larger than differences with offline registers. In terms of *of*-phrase and *for*-phrase phrasal postmodification specifically, the proportions of phrasal postmodification are generally higher in online registers compared to offline, and the effect-size differences are larger and more distinct between these groups, especially where the rate of occurrence is low. This concludes phrasal postmodification as densification device.

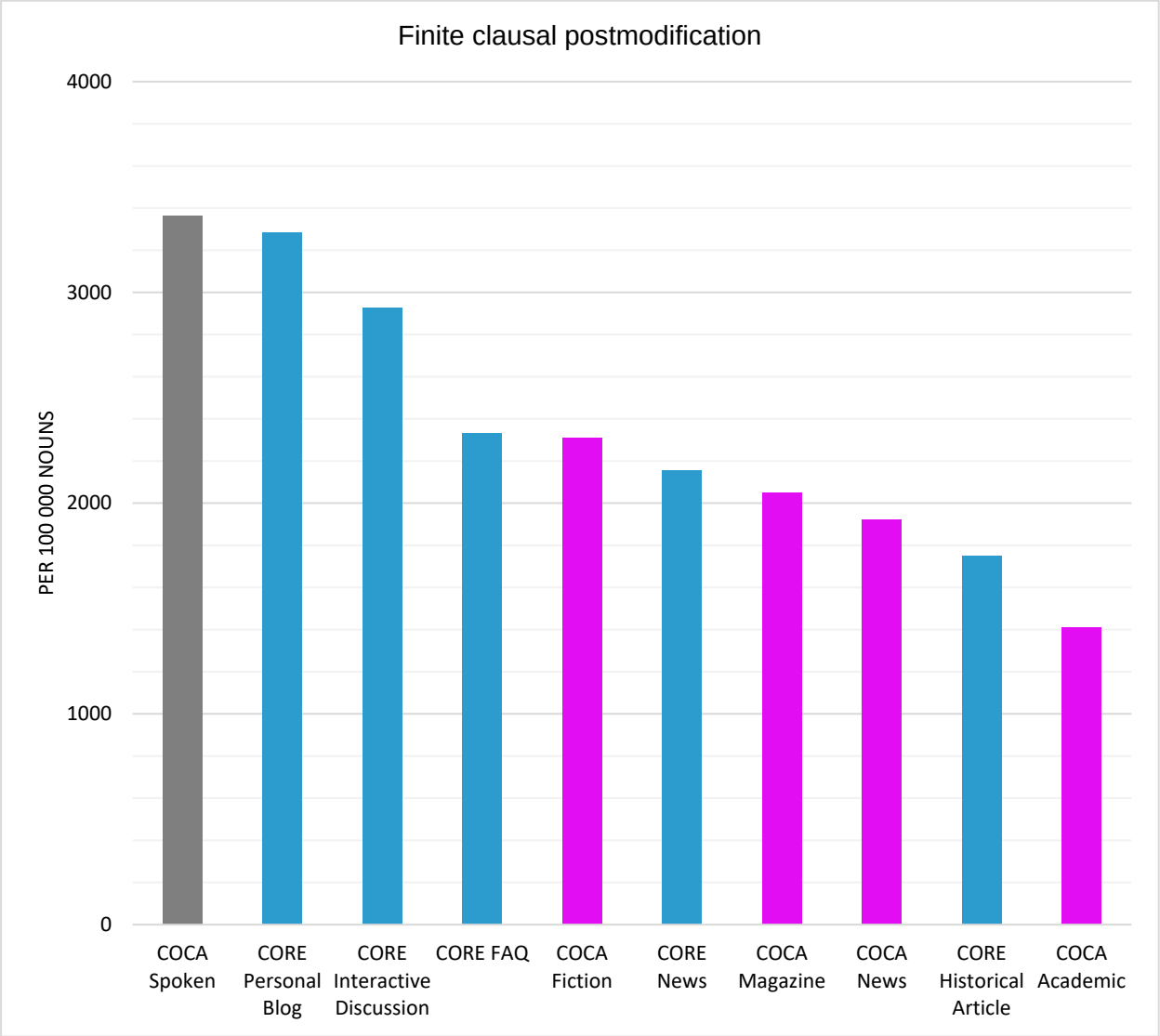
4.4 Postnominal clausal modifiers

The following section describes density patterns of postnominal clausal modification in offline and online registers. Clausal postmodification is less dense than phrasal postmodification in the sense that the modifying structure takes on the form of a clause, although still embedded within the noun phrase. Density can further be distinguished in terms of finite and nonfinite clausal postmodification. The *LGSWE* shows that the written registers, such as news and fiction, have higher frequencies of finite clausal postmodification than conversation does (Biber *et al*., 1999:606–610). Of the written registers, academic writing has the lowest frequency. Finite relative clauses are relatively uncommon in conversation compared to these other written registers. Further, nonfinite postmodification is less common than finite relative clause postmodifiers, but they too are also rare in conversation compared to written news and academic text types (Biber *et al*., 1999:606).

4.4.1 Finite relative clauses

Figure 4-3 sets out the total combined relative frequencies per 100 000 nouns for *that*- and *which*-clauses, unmarked relatives and relative pronouns and adverbs introducing relative clauses. The registers are ranked from the highest to lowest normed frequency. Figure 4-3 below presents finite relative clause postmodifiers, with the registers ranked from highest to lowest normed frequencies per 100 000 nouns in terms of total finite clause constructions. Nonfinite clauses are ranked later in a separate graph (Figure 4-4). The same colour scheme is applied to distinguish online from offline registers, where online written registers are marked in teal, offline written in purple, and offline spoken in grey.

Figure 4-3: Finite postmodifying clauses per 100 000 nouns



The data from Figure 4-3 show slightly contradictory patterns compared to Biber *et al*. (1999:606). As is evident from the register rankings in this figure, Spoken has the highest overall relative frequency of clausal postmodification across all registers and is clearly distinguished from some written registers such as offline News, Historical article and Academic. The distinction between Spoken and written online Personal blog and Interactive discussion is less obvious. The examples from concordance data below illustrate the use of finite clausal postmodification in Spoken and Personal blog (9–11) compared to those in Academic (12–13):

9. what I've seen so far is a team of very dedicated **engineers** that are pulling together a pretty wide array of disciplines. (Spoken, COCA)

10. I consider to be the most selfless **thing** that she could have possibly done for me (Personal blog, CORE)

11. So much of the music I like aims to conjure **emotions** that are complicated and heavy, which is challenging and fascinating and great. (Personal blog, CORE)

12. the **studies** that employ growth mixture modeling, focused on identifying latent classes of growth patterns (Academic, COCA)

13. Students are also given specific **criteria** that should be included in their composition (Academic, COCA)

Of the written registers, online involved Personal blog and Interactive discussion have the highest frequencies. Frequencies sharply decline from FAQ and it is the expository and information-driven writing, as well as offline News, that have the lower frequency scores of clausal postmodification. Considering the register groups COCA and CORE, a cline from informal to formal exists in descending order for each corpus group. A similar observation is made with regards to the CORE texts, with involved writing (Personal blog and Interactive discussion) at the higher-frequency range and informational writing registers (FAQ, online News and Historical article) on the lower frequency end.

Further, it is mostly offline registers that appear on the lower-frequency range, whereas online registers are more grouped at the higher-frequency end. Even FAQ, which had higher proportions of phrasal premodification and postmodification, retains some of the higher relative frequencies of clausal postmodification.

Of the structures making up the data in Figure 4-3, two showed unexpected patterns. These are *that*-clauses and *which*-clauses introducing postmodifiers. According to Biber *et al*. (1999:606–611) *that*-relative clauses are more common in written registers; they are the most common in fiction and have the lowest relative frequency in conversation. Postmodifying *which*-relative clauses are the most common in academic writing, somewhat less common in news and fiction and very uncommon in conversation (Biber *et al*., 1999:609–611). The following descriptions below address some register similarities and contrasts with regard to *that*-clauses (first) and *which*-clauses (afterward).

While a cline from involved to informational purpose is noted for online registers, this is not the case for offline registers. Additionally, *that*-clauses, as a density device, do not distinguish the offline and online registers in any obvious way. The offline and online registers are rather mixed when ranked from highest to lowest relative frequency, and most written registers have similar relative frequencies, except for the outermost contrasts formed between written registers and Spoken (with the highest relative frequency) and with Historical article (with the lowest frequency), respectively. This is reflected in the data with regard to similarities among register pairs in Table

4-9 below. This table presents the effect size difference of the most similar register pairs for *that*-relative clause postmodifiers across all registers. The register pairs appear in the first column, where online registers are marked in teal, offline registers appear in purple. The effect size differences for each pair are listed in the second column and are sorted from the most to the least similar. (See Table A.7 in appendix for complete comparison.)

Table 4-9: Most similar register pairs for *that*-relative clauses

Register pairs	%DIFF
News_FAQ	0,00
Magazine_Interactive Discussion	0,10
CORE News_Fiction	0,67
Academic_CORE News	5,26
Academic_Fiction	5,96
Interactive Discussion_News	6,32
Interactive Discussion_FAQ	6,33
Magazine_News	6,43
Magazine_FAQ	6,44
FAQ_Academic	10,37

The nine most similar pairs in Table 4-9 differ less than 10%. The top three pairs evidently show no difference between registers in the respective pairs, but all of the differences listed in this table are considered small and are practically similar in their use of *that*-clause postmodification. The registers most commonly involved in these smaller differences are offline News, Magazine and Academic in COCA, and FAQ, online News and Interactive discussion in CORE. Of these pairs, online–offline similarities are more common than interregister pairs with a small effect size difference. These pairs also show that there is little distinction in terms of register type. For example, the pairs Academic and online News, Academic and Fiction, and Interactive discussion and FAQ, are not distinguished in their use of *that*-clause postmodifiers.

Further, Table 4-10 below lists the most different register pairs for *which*-relative clauses for all pairwise comparisons. To provide some context to the data in this table, brief reference is made to the distribution patterns of *which*-clausal postmodifiers here. First, all five online registers have higher relative frequencies compared to offline registers, which all fall within the bottom range. In fact, all finite clausal postmodifiers except unmarked relatives lead to more CORE registers

appearing in the higher-frequency range than the offline COCA registers. In the online registers, *which*-clauses are most common in Historical article and the least common in online News; in offline registers, they are most common in Spoken, followed by Academic.

Which-clauses are very uncommon in the offline registers compared to the online registers, and this is reflected in the large frequency differences in Table 4-10, which mostly involve online–offline contrasts within register pairs. The register in the left position in each pair has the higher relative frequency, which in this case is always an online register clearly demarcated in teal. Offline registers are marked in purple and all appear here on the right of each pair. The percentage difference score is shown in the second column, and these are sorted from the most to the least similar. The ratio of relative frequency value appears in the last column. (See Table A.8 in Appendix.)

Table 4-10: Most different register pairs for *which*-relative clauses

Register pairs	%DIFF	RRF
Historical Article_News	3 085,67	31,86
Historical Article_Magazine	2 271,19	23,71
FAQ_News	2 211,40	23,22
Interactive Discussion_News	1 766,29	18,66
FAQ_Magazine	1 620,45	17,20
Personal Blog_News	1 469,88	15,70
Historical Article_Fiction	1 298,46	13,98
Interactive Discussion_Magazine	1 289,13	13,89
CORE News_News	1 284,83	13,85
Personal Blog_Magazine	1 068,51	11,69

The effect size differences reported for *which*-clause postmodifiers are some of the largest among all modification and complementation types. While *which*-clauses are very infrequent in the registers, (the second most infrequent clausal postmodifier type among all finite postmodification), they create some of the largest differences between online and offline registers. All the differences in this table indicate that the largest differences exist between CORE and COCA registers and are not interregister-based.

Evidently, the online information registers Historical article and FAQ more commonly contrast with offline narrative (Fiction) and nonnarrative (offline News and Magazine) writing than other written online registers. Even involved Interactive discussion contrasts more with offline News and Magazine than online News to offline News.

4.4.2 Nonfinite clauses

This section describes the data patterns related to nonfinite clausal postmodifiers. Nonfinite postmodification is structurally more dense than finite clausal postmodification. Nonfinite clauses are generally more common in written nonfiction registers and the least frequent in conversation (Biber *et al.*, 1999:606). Postmodifying *to*-clauses are equally rare across all four registers of the *LGSWE*, but most common in fiction, and they have slightly lower frequencies in news and in academic writing. They are the least frequent in conversation (Biber *et al.*, 1999:606). Both participial *ed*/*en*-clauses and *ing*-clauses are more common in academic writing and news than in fiction and conversation. *Ing*-clauses are also more common in fiction than *ed/en*-participial clauses (Biber *et al.*, 1999:606).

Figure 4-4 below presents the three individual nonfinite postmodifiers in the CORE and COCA registers. The online and offline written registers are distinguished by colour codes (offline written registers are purple, online registers are teal and Spoken is grey). Lighter and darker tone bands distinguish the individual nonfinite density features (dark signifies *ing*-clauses, light signifies *ed/en*-clauses and medium dark signifies *to*-clauses). These tone bands are stacked in columns for each online and offline register. Registers are ranked according to the total combined relative frequency of each individual feature. Normed frequencies occur per 100 000 nouns.

Figure 4-4: Types of nonfinite modifying clauses per 100 000 nouns

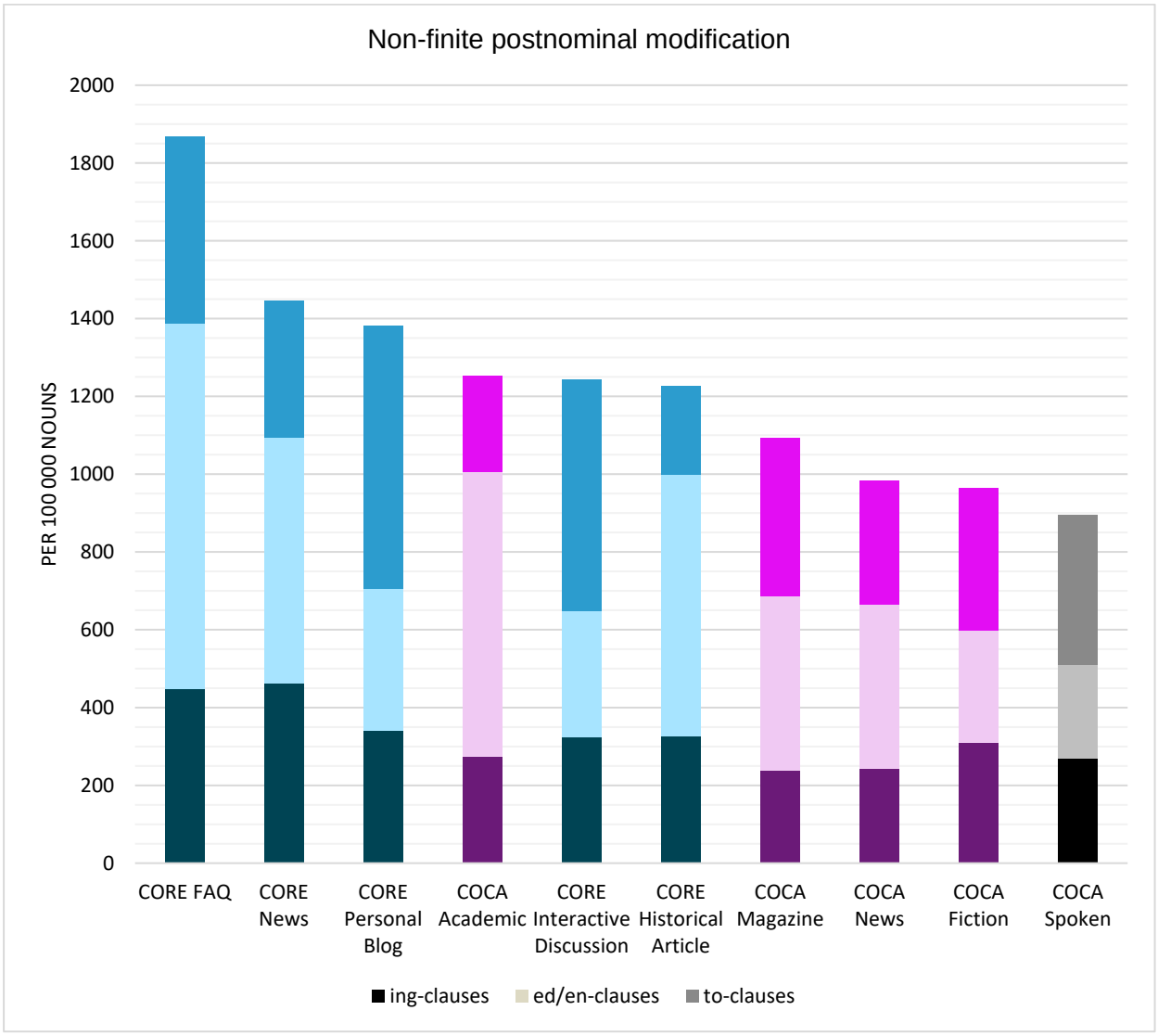


Figure 4-4 first shows that Spoken has the lowest overall relative frequency of nonfinite postmodification, where the online FAQ has the highest proportion of combined nonfinite density features. Viewing the offline registers (purple) in isolation indicates some formal-to-informal cline exists in relation to Spoken. The position of Historical article in relation to other online registers makes such a cline less prominent for registers in CORE, although FAQ has the highest relative frequency within this group. Spoken has higher frequencies of *to*-clauses than would be expected, even more so than Academic; *ing*-clauses are also not the least common in Spoken, but they are in written Magazine and offline News. Spoken does however have the lowest frequency of *ed/en*-clauses, which mimics patterns in the *LGSWE*.

Considering all registers, there is no discernible pattern with regard to text type or purpose for most registers between the outermost registers (FAQ and Spoken). For example, involved and popular writing, such as Interactive discussion, and expository Academic are similar. Clearly,

ed/en-clauses mark Academic more than they do Interactive discussion, and *to*-clauses are more common in Interactive discussion than in Academic. Academic, on the other hand, has lower uses of *ing*-clauses and *to*-clauses, but markedly increased uses of *ed*/*en*-clauses compared to both Interactive discussion and Personal blog.

Registers rank differently depending on the clausal modifier type. *Ing*-clauses are more or less equally common in both online and offline registers. There is a clear split between online and offline registers here, where online registers (CORE) are grouped together on the higher relative frequency range, and offline COCA registers bundle on the lower-frequency range. There is no discernible pattern with regard to register purpose in either offline or online registers. Spoken is also not distinct from other written registers in its use of *ing*-clauses; Magazine has the fewest occurrences here.

Frequencies of *ed/en*-clause postmodifiers are more varied across online and offline registers. In this case, Spoken has the lowest frequency and FAQ has the highest relative frequency. *Ed/en*-clauses are also very common in informational registers, including Academic, Historical article and online News. Fiction and involved writing generally have lower relative frequencies. Despite some pattern emerging in terms of register purpose, more online registers appear in the higher-frequency range than offline registers do. In addition, the online registers in the bottom group (Personal blog and Interactive discussion) have higher proportions of *ed/en*-clausal postmodification than offline Fiction and Spoken.

To-clause postmodification makes no clear distinction between register purpose; however, it does place more online registers – Personal blog, Interactive discussion and FAQ (in this order) – in a higher-frequency range. Online Historical article has the lowest frequency of all the registers; News (CORE) also falls in the lower-frequency group. However, most offline registers, including Fiction, offline News and Academic have lower proportions of *to*-clause postmodification. It is only offline Magazine and Spoken that have slightly higher occurrences; Spoken is therefore also not obviously different from written registers in this case.

So far, the data for nonfinite clausal postmodification show that density is more pronounced in online than offline registers. While *ing*-clauses more clearly separate online and offline registers from one another than *to*-clauses and *ed/en*-clauses, the differences between them are less pronounced than those created by *to*-clauses and especially *ed/en*-clauses. Table 4-11 lists the most different nonfinite postmodifying clauses sorted from largest to smallest percentage difference scores. Online registers are marked in teal, offline registers in purple, and Spoken in grey. (See Table A.9 and Table A.10 for complete list of pairwise comparisons for each construction.)

Table 4-11: Most different nonfinite postmodifying clauses

Clause type	Register pairs	%DIFF	RRF
ed/en-clause	FAQ_Spoken	291,53	3,92
ed/en-clause	FAQ_Fiction	225,18	3,25
ed/en-clause	Academic_Spoken	204,57	3,05
to-clause	Personal Blog_Historical Article	197,41	2,97
ed/en-clause	FAQ_Interactive Discussion	189,25	2,89
ed/en-clause	Historical Article_Spoken	180,56	2,81
to-clause	Personal Blog_Academic	173,47	2,73
ed/en-clause	CORE News_Spoken	163,06	2,63
to-clause	Interactive Discussion_Historical Article	162,03	2,62
ed/en-clause	FAQ_Personal Blog	157,51	2,58

What Table 4-11 evidently shows is that differences in *ed/en*-clauses are more pronounced between register types, rather than the offline–online mode. Where offline–online differences occur, they seem to contrast register type in terms of purpose. Specialist writing and information-driven registers clearly contrast with popular writing. Of all nonfinite postmodification types, 21 pairs differ with a percentage difference score of 100 or more. Of these 21 pairs, 15 are differences involving *ed/en*-clauses. *Ed/en*-clause postmodifiers in informational texts (14–15) show a contrast in terms of their use in involved writing (16–17) such as Personal blog and Interactive discussion:

14. The per person consumption of **carrots** used for freezing and canning in 2009 was 1.4 and 0.8, respectively (Historical article, CORE)

15. The **adjustment** required is well-documented in Dostoevsky's second to last novel (Historical article, CORE)

16. It will make a difference if I buy toilet **paper** made from recycled paper, and I can get that at Walmart (Personal blog, CORE)

17. I am writing this complaint with regard to my **grievance** caused by your airline company (Interactive discussion, CORE)

In fact, concordance data for Interactive discussion show a general trend that when *ed/en*-clausal postmodifiers as in (17) are used, the head nouns are sometimes abstract: *presentation*, *opinions*, *risks*, *issues*, *errors*, *results, skills, capabilities* and *decision*.

Larger differences pertaining to *to*-clauses more commonly involve seemingly involved written contrasts to informational registers, for instance Personal blog and Historical article, Personal blog and Academic, and Interactive discussion and Historical article. In all of these pairs, it seems that the online mode places fewer restrictions on the author since there are no obvious spatial restrictions, and the information in the examples from online FAQ (18) and Personal blog (19–20) could have been formulated in a way that avoids the modifier type altogether:

18. What are the **steps** to apply for a grant? #You first needed to register each property and receive (FAQ, CORE)

19. Do I have an **appointment** to go to, nope. Do I have a **friend** to meet, no. It's not really anything particularly 'time sensitive' (Personal blog, CORE)

20. Now if I can only find a **way** to squeeze 6 more hours out of my day without totally neglecting my family ... (Personal blog, CORE)

Nonfinite postmodification is very characteristic of online registers rather than of offline registers; however, individual postmodification types rank registers differently. *Ing*-clause postmodification proportions are most evident in terms of establishing an online–offline difference, with all online registers having overall higher proportions; frequencies based on *ed/en*-clauses and *to*-clauses also generally place online registers in the higher-frequency range, but not as clearly as *ing*-clause postmodifiers do. *Ed/en*-clauses are also the only nonfinite postmodifier generally showing a formal–informal progression across all registers.

4.4.3 Postnominal modifying clauses introduced by a preposition

The following section provides a brief description of postnominal modifying clauses that are introduced by a preposition. Since the phrasal–clausal distinction in modification is important to density (phrasal structures being structurally more dense than clausal ones), the data patterns surrounding postnominal modifying clauses introduced by a preposition are treated separately from phrasal postmodification and also from finite and nonfinite clausal postmodification.

Postnominal modifying clauses in the concordance data typically include *ing*-clauses, *wh*-clauses as well as quotative uses:

21. didn't want to get desperate and tweet **stuff** like, 'Uhhh, breaking your stick in the middle of a breakaway in OT' (Personal blog, CORE)

22. making money from people searching for **things** like 'download we are young fun.' (Personal blog, CORE)

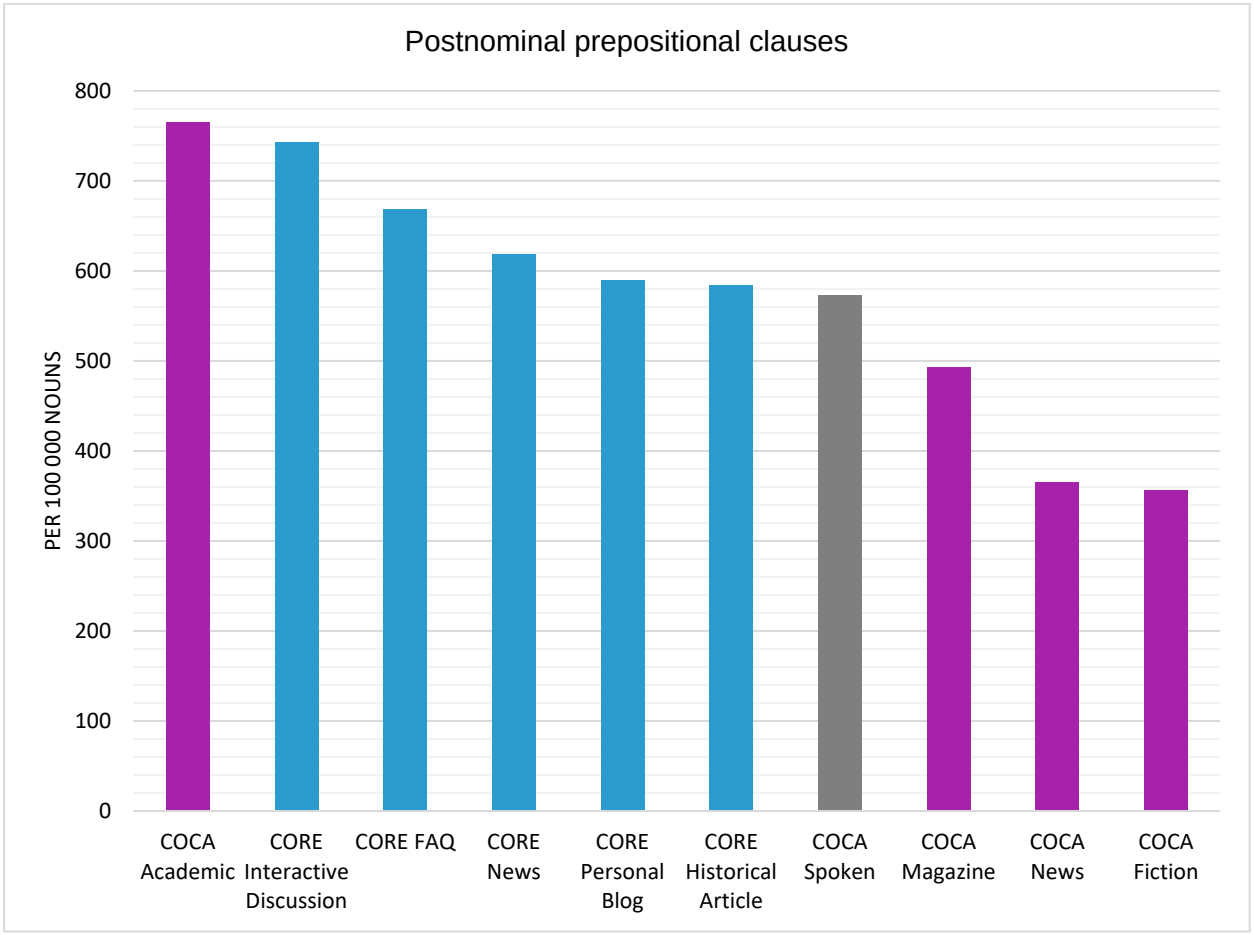
23. It includes **instructions** on how to avoid disturbing bears and what to do if attacked. (News, CORE)

24. The **idea** of putting a further strain on motorists is unfair. (News, CORE)

25. Then there was the **flat** in which there was a (shaver's?) switch in the bathroom (Interactive discussion, CORE)

These structures are less dense than phrasal postmodifiers and are therefore treated separately from phrasal postmodification. Their relative frequencies per 100 000 nouns are presented in Figure 4-5 below. Colour codes distinguish between online writing (teal), offline registers (purple) and Spoken (grey).

Figure 4-5: Postnominal prepositional clauses per 100 000 nouns



Firstly, the relative frequencies presented in Figure 4-5 show that postnominal clausal modifiers following a preposition are very uncommon in both the online and offline registers compared to other modifier types.

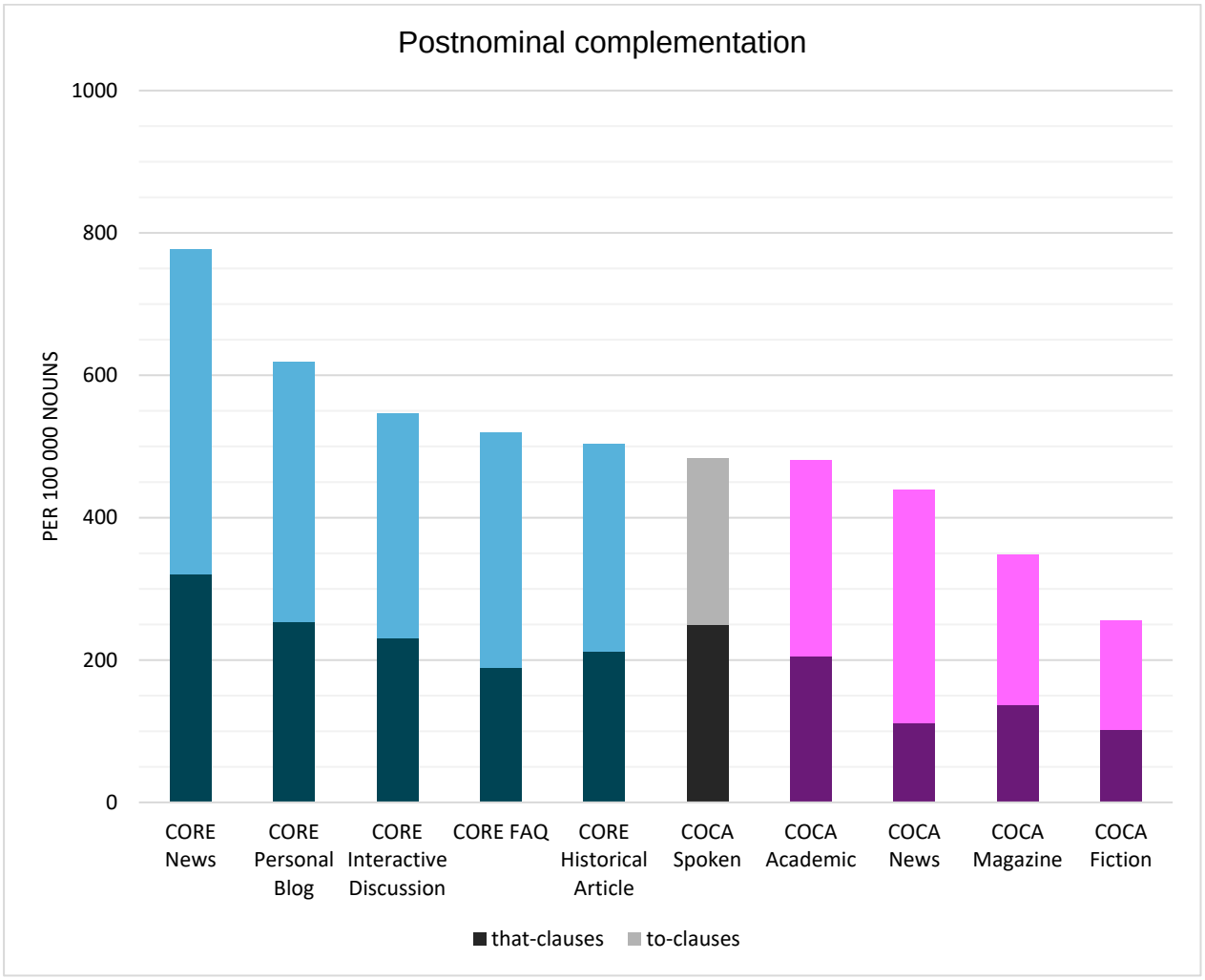
Relative frequency in Spoken is not obviously different from other registers, specifically written Historical article, Personal blog and online News. The lowest relative frequency is found in Fiction and offline News. The highest frequency is Academic, a contrast with offline News, but this structure does not differentiate between most of the popular written texts or the spoken component. While more online written registers have higher frequencies than offline texts, there seems to be no real variation between these texts.

4.5 Postnominal complement clauses

This final section describes the proportions of noun complementation as a density device. Structurally, noun complement clauses are denser than clausal postmodifiers. Further, findings from the *LGSWE* indicate that noun complementation is not as common as phrasal postmodification and that the two major complement clauses, *to*-clauses and *that*-clauses, are more common in written registers than in conversation; *to*-clauses being especially common in news and *that*-complement clauses being the most frequent in academic writing (Biber *et al*., 1999:647).

Figure 4-6 below presents the normed frequencies for noun complement *that*- and *to*-clauses per 100 000 nouns. The online registers are marked in teal, offline registers in purple and Spoken in grey. The two complement types are stacked and differentiated through tone bands. (*That*-complement clauses are dark and *to*-complement clauses are light). Registers are ranked from highest to lowest relative frequency based on the total combined frequency for complementation overall.

Figure 4-6: Noun complement clauses per 100 000 nouns



First, noun complementation is infrequent compared to the other modifier types described so far. In terms of overall relative frequency and individual complementation types, online News has the highest proportion across all registers. Fiction in COCA has the lowest relative frequency, also for both types of complement. The patterns here also show that Academic has more *to*-complement clauses than *that*-complement clauses, where both offline and online News have a more distinguished difference in the types that coincide with findings in the *LGSWE*.

An additional contrasting finding is that Spoken does not have the lowest relative frequency of either feature. It has the third-highest relative frequency in *that*-complement use, although it does have the third-lowest use of *to*-complement clauses. While Spoken will likely differ from the outermost registers such as Fiction, Magazine and online News, it is evident from this figure that it is similar to Academic and Historical article at least, and therefore not overtly different from written registers. Spoken contains sections of texts that report or relay information on events or specific topics. The excerpts from Spoken seem to illustrate that when noun complement clauses are used, these are in a similar lexical context as that of reportage in news:

Text 5:

I don't know if it's so much cool. I mean, it's definitely – it seems to be an ***attempt*** *to get attention. I presume it's considered cool. They are traded amongst, you know, numerous students once they're released out there. And again, when the student is unconcerned about what they've just done, I have to presume it is some kind of a status* ***symbol*** *or an* ***intent*** *to get attention in some form* (Spoken, COCA)

Text 6:

…and we all thought we have a new president, we have a strong, clear science statement, we have a ***mandate*** *to act, and of course, it was within weeks that the U.S. walked from the Kyoto Protocol…* (Spoken, COCA)

Text 7:

Our case was premised on the ***fact*** *that Todd Sommer was poisoned by arsenic. And when the information came forth from experts that are renowned, that say they can't tell us with a medical certainty that he was, in fact, poisoned by arsenic, at that point, we stop. That is the bottom line. We now have a concern, we now have reasonable doubt, we now have a* ***duty*** *to dismiss the case* (Spoken, COCA)

The abovementioned excerpts can be compared to News (COCA):

Text 8:

I think it's Ross ***responsibility*** *to fight and stay in office. He was elected by the people, and if the people want to remove him, that's their call* (News, COCA)

Text 9:

'As an artist you have a ***responsibility*** *to be relevant, to put something out there that has meaning,' said Denver hip-hop artist Molina Speaks…Art has the* ***power*** *to radically reframe the conversation. 'What's new is new again* (News, COCA)

Further, the colour-coded columns immediately indicate that the online registers are separated from offline registers. Sorting according to individual complementation types led to the same finding, namely that most online registers have higher relative frequencies of *to*-complement clauses (News, Personal blog, FAQ and Interactive discussion) and *that*-complement clauses (News, Personal blog, Interactive discussion and Historical article). This leaves most offline

registers in a lower-frequency range. Neither group (CORE or COCA) indicates any pattern based on register purpose.

Not only is noun complementation more common in online writing, but the differences between online and offline are rather large. *That*-complement clauses have more register pairs with a percentage difference score of over 100,00. Table 4-12 below therefore sets out the effect size differences for the most different noun *that*-complement clauses. Online registers are marked in teal and offline written registers are marked in purple. Spoken appears in grey. (See Table A.11 in Appendix).

Table 4-12: Most different *that*-clauses as noun complement

Register pairs	%DIFF score	RRF
CORE News_Fiction	214,15	3,14
CORE News_News	188,54	2,89
Personal Blog_Fiction	147,80	2,48
Spoken_Fiction	144,88	2,45
CORE News_Magazine	134,09	2,34
Personal Blog_News	127,60	2,28
Interactive Discussion_Fiction	125,66	2,26
Spoken_News	124,91	2,25
Interactive Discussion_News	107,26	2,07
Historical Article_Fiction	107,06	2,07

All the differences listed in Table 4-12 show that *that*-complement clauses are more than twice as common in the online registers than in the offline texts. The largest differences often involve online News compared to offline Fiction, online News and Magazine. It is also these three offline registers that repeatedly feature as the register with the lower proportion of *that*-complement clauses.

It was stated earlier that Spoken is more similar to most written registers, such as Academic and Historical article (as is evident from Figure 4-6). As with most other modifier types previously discussed, Spoken does form larger contrasts in terms of relative frequency with the outermost registers on the scale. Evidently, Spoken differs the most from written Fiction and offline News respectively, having more than twice as much *that*-complement clauses than the written offline

registers here. Apart from the difference between online News and offline News being the second largest (as shown in this table), this difference is probably the most pronounced in terms of *that*-complementation, despite the infrequent occurrence of noun complementation clauses in this data set.

4.6 Conclusion

This chapter presented the relative frequencies of density features associated with noun phrase compression in terms of differences between a spoken register and nine other written registers of English. These same registers were also compared in terms of online (CORE) and offline (COCA) registers to establish which registers have the higher and lower frequencies of noun phrase density devices. The differences in the ranking of the 10 registers were further described in terms of the size of the differences between the registers for many of the modifying and complementising structures.

Comparing online and offline registers in terms of density has revealed interesting patterns relating to register types or text purpose. More importantly, however, it has been shown that online registers are characterised by features of information compression in proportions that often exceed that of offline registers. At times, information density even characterises online involved and popular or general writing on the web to an extent which is unexpected and contrasts with patterns of traditional findings.

Proportions of phrasal premodification are, overall, higher in offline registers than online registers, but the effect size of the differences between the middle point where offline diverges from online registers is small. In terms of premodifier type, online FAQ has the highest frequency, but overall offline registers still have the higher overall frequencies with more offline registers appearing in the higher range. Online registers have higher normed frequencies for both proper noun and adjective premodifiers (even though Academic has the highest). It seems that online registers are less marked in terms of common noun premodifiers than the other two types.

Phrasal postmodification is more pronounced in online registers in terms of overall combined frequencies and *of*-phrases specifically. Offline Academic has the highest relative frequencies of *in*-phrase and *for*-phrase postmodifiers, and these postmodifiers are more common in offline writing than online registers. Phrasal postmodification also shows two of the largest inverse patterns for offline News and online Historical article compared to their premodification uses. Offline News avoids the use of phrasal postmodifiers to the extent that it has the lowest use of combined phrasal postmodifiers and specifically *of*-phrases. However, this is not the case with *in*-phrases and *for*-phrases. Offline News has some of the highest uses of *in*-phrase postmodification

in the data set. Its use of *for*-phrases is less common, but not as infrequent as *of*-phrases. This suggests restrictions related to the offline mode, where space limitations may prevent unnecessary elaboration, which causes News to avoid postmodifiers (at least phrasal postmodifiers) and to employ the alternative *s*-genitive instead. This possibility could be further explored by counting and comparing *s*-genitives with the *of*-possessive counterpart in offline News. Additionally, online News has higher proportions of phrasal postmodifiers and might serve as a control group. As shown in the data, offline News also has some of the lower proportions of noun clausal postmodification and complementation, which furthers the notion that it functions under the strain of text length.

In terms of clausal postmodification, distribution scores show that the difference between offline and online registers is even more pronounced than that of phrasal premodification and phrasal postmodification. Although, involved writing (Personal blog and Interactive discussion) and Spoken have higher frequencies of finite clausal postmodifiers, which is unexpected given that written fiction and information-driven registers usually have the higher occurrences of these modifiers overall. (Some finite postmodifying clauses are more frequent in some registers depending on the relative marker.) The higher occurrences in Spoken may suggest that speakers more readily have recourse to clausal modification devices, which are less complex than, for instance, phrasal modifier types, when specifying referents in a speech situation.

Postnominal clauses introduced by prepositions, being very infrequent compared to other types, show a clear online–offline divide (second to overall complementation patterns). With the exception of Academic, which has the highest overall frequency, all other online registers are adjoined in the higher-frequency range. Interestingly, Interactive discussion has the highest use in clauses introduced by *of*.

Combined nominal complementation frequencies show the starkest contrast between online and offline registers. Individually, *that*-complement clauses and *to*-complement clauses also place more online registers in the higher-frequency range. *That*-complementation, which marks expository writing, is especially common in online News and Personal blog. Spoken also has higher occurrences of this structure (all surpassing Academic). The online mode may accommodate more complex structures in a different range of text types, since, again, length restrictions may be less important in blog sites and online information pages. *To*-complementation is also most common in online News and Personal blog, but is also common in offline News. Complement clauses are structurally more dense than clausal postmodifiers, and offline News, on the other hand, seems to opt for these structures rather than less dense forms.

CHAPTER 5 CONCLUSION

5.1 Introduction

This chapter offers a synthesis and conclusions about the way in which selected online and offline registers employ grammatical features associated with density in the noun phrase. Three levels of comparison were used to answer the research question. The first focused on comparisons between the spoken component and written registers and the second one on online and offline registers. The third level compared the proportions of modification and complementation features in terms of 45 pairwise comparisons using effect size as measure. Density, which involves modification and postnoun complementation patterns, encompasses a small set of defined features that are specific to noun-headed phrases. The extent to which noun heads are modified in discourse depends on a complex and varied range of factors underlying the situational context. Discourse structure is ultimately shaped through these factors, and this includes the role of a web-based interface. Further, the discourse styles associated with noun phrase density are often registers that have an information-driven purpose. Text types or register purpose vary, and this is evident in different linguistic characterisations of those discourses. Noun phrase density is primarily a common feature of written language, but certain less complex modification types, such as finite relative clauses, are also very prevalent in spoken language. This testifies to the ability of speakers to present noun referents to an audience using this type of grammatical resource.

5.2 Chapter summary

Chapter 1 contextualised this study in terms of noun phrase density with reference to other studies that have focused on variation among registers in terms of these features specifically. Internet-based language has received ample attention from various areas of linguistic inquiry. Aspects surrounding grammatical variation and pragmatics have been topics of investigation with regard to web-based language use. While MD analyses have been performed on the CORE registers, little else has been done in terms of register variation of online written English. In light of this and the role of nominal modification as an identifier of dense language, the research question could be framed accordingly and applied to web-based written registers

The factor analyses from Biber and Egbert (2016) and descriptions of registers on the corpus interface were the only reference for expectations regarding differences in frequencies of modification and complementation in the noun phrase for these specific selected registers. Further, descriptions of noun phrase modification patterns from the *LGSWE* (Biber *et al*., 1999) were used to contextualise and compare selected registers from CORE and COCA. Based on this, it was expected that a higher proportion of dense noun phrases would be present in

informational written registers. It is against this backdrop that the frequencies were compared and interpreted qualitatively.

Chapter 2 provided the grammatical framework used to construct search strings for each modifying and complementising feature. Two reference grammars informed the choice of modification and complementation constructions in the noun phrase to include in this study: Huddleston and Pullum (2002) and Biber *et al*., (1999). These grammars also aided in formulating practicable search strings with which to draw concordances. In sum, a defined group of features was selected prior to the investigation in order to establish the degree to which they characterise these 10 registers in terms of density, with the aim of comparing and contrasting online and offline registers as a group, as well as registers with each other, regarding the usage of each of the following typical noun phrase elements: nouns as premodifiers, adjectives as premodifiers, postnominal prepositional phrases, postnominal modifying clauses and postnominal complements.

A great deal of care was taken with the method to ensure that the data were as reliable as possible. While the surface structure was the main point of departure in creating a search string and searching and counting grammatical features, concordances were manually filtered to establish a good precision rate for each structure. The search strings yielded concordance data containing tag errors and other syntactic functions not relevant to density. Further, for most features, two random concordances were drawn per feature in order to calculate the average precision rate of each search string. These precision rates were then used to compute a more accurate estimated frequency for each grammatical feature. After a time-consuming process, the data structure allowed a quantitative analysis involving the comparison of normed frequencies across registers (using a noun count as normalisation base) and the inclusion of a measure of effect size. The method mainly relied on percentage difference as a quantitative measure of the difference between the normalised scores of any two registers in the pairwise comparison. This measure revealed information about both the differences and similarities between online and offline registers.

Chapter 4 provided a quantitative description of noun phrase modification and complementation patterns supported by a qualitative interpretation of these results. The chapter focused on premodification, phrasal postmodification, clausal postmodification and finally, noun complementation patterns.

In terms of nominal premodification, two major distinctions were made with regard to modifier type: nouns and adjectives (nouns included proper nouns as premodifiers). Generally, the frequency data of premodifiers largely concurred with findings that show registers with

information-driven purposes and specialist writing have higher proportions of premodification than involved and general writing. However, at times even the online involved registers had higher proportions of premodification than Fiction and Historical article. When considering premodifier types individually, adjective premodification distinguishes online registers from offline registers more clearly, with more online registers containing higher proportions than offline texts (even though Academic had the highest relative frequency among all registers here). This is not the case with premodifying nouns. In this case, the proportions of premodification are the highest in online FAQ and high in online News, but the remaining online registers (Interactive discussion, Personal blog and Historical article) have proportions in a lower region, with Historical article having the lowest. The relative frequencies of proper nouns as premodification device showed no obvious pattern in terms of register type, but online registers had slightly higher proportions than most offline registers. In some individual premodification patterns, density does distinguish online from offline registers to some extent, but it seems that these distinctions might be related to register type rather than the online mode. In addition, these differences are not nearly as pronounced as the differences in terms of phrasal postmodification.

Nominal phrasal postmodification made sharper distinctions in terms of the rankings of offline and online registers, and density therefore seems to be more important to the characterisation of online writing compared to offline registers. In terms of combined phrasal postmodification devices, online Historical article had the highest proportional use. Additionally, online News, FAQ and Personal blog (an involved register type) all had higher relative frequencies than COCA Magazine, Spoken, Fiction and News. Academic is the only register in the higher-frequency range, probably due to it being expository in nature. However, not all prepositions introducing phrasal postmodifiers make equal distinctions in this way. *Of*-phrases and *for*-phrases were most pronounced in separating online from offline writing. Both these structures, as well as *other*-phrases, showed more online registers adjoining in the higher-frequency range. Where an offline register did appear in the higher range it was often Academic and not any of the other offline written texts, including News (COCA). In fact, offline News had the lowest relative frequency of combined phrasal postmodification. It was continuously surpassed by the online variant in most of the phrasal postmodifier types, but the most pronounced difference here between the two news types was their use of *of*-phrases (they are otherwise more similar than different).

Finite clausal postmodifiers, which are less dense than phrasal modifiers, were the least common in expository writing (Academic), nonfictional narrative writing (Historical article) and offline News. They were most common in Spoken, followed by online involved written registers (Personal blog and Interactive discussion). Nonfinite clausal postmodifiers showed no obvious rank in terms of register purpose, but online registers overall had higher frequencies of nonfinite structures, where

offline registers had the lower proportions. Spoken (having the highest relative frequency of finite postmodification) and written involved registers had higher proportions of finite clausal postmodification, which was not expected given the distribution data of finite postmodification patterns in the *LGSWE*. One particular structure that stood out is *which*-clause postmodifiers. They were very infrequent in offline registers and exponentially more common (as per percentage-difference scores) in all online registers. It was also surprising that Spoken had higher frequencies of *which*-clauses than Academic, and by extension all other written offline registers in COCA. Nonfinite postmodification, in turn, ranked Spoken with general fiction and nonfiction (Fiction, Magazine, offline News) in the lower proportions, where FAQ, online News and even Personal blog (which surpassed Academic) had the highest relative frequencies. Taking into consideration that individual patterns are not distributed similarly across registers, nonfinite postmodification generally distinguished online writing from offline registers, where online registers indicated higher relative frequencies.

Clausal postmodifiers introduced by prepositions were uncommon in all registers, but apart from Academic, which had the highest proportion of combined postmodifying clauses, all online registers adjoin in the remaining slots from higher to lower relative frequency, leaving the residual offline registers in lower frequency positions. Further, these structures are the second-most frequent in Interactive discussion, an involved written register. When individual prepositions were analysed (with clauses introduced by *of* being the most frequent among all 12 prepositions), Interactive discussion contained the highest relative frequency of *of* + clause structures, where its frequency is more than one and half times as frequent as that of Academic. Academic, on the other hand, has more than twice the number in postmodifying clauses introduced by *in* and *for*. Evidently, different types of modifier lead to differences in the rankings of registers, and not all features are equal in terms of the size of differences between normed scores. Therefore, while these features are traditionally associated with informational writing, they also frequently occur in seemingly more general writing, although the differences are not large.

Finally, density was perhaps most pronounced in terms of noun complementation, where all online registers grouped in the higher-frequency range (more so than offline registers). However, as with all modifier types, individual structures create variation in the rankings of these registers. Still, online registers were more clearly marked for noun phrase density in this regard than offline writing. In addition, most of the larger contrasts in terms of *that*-complement clauses were between involved online registers (Interactive discussion) and popular or general written offline registers such as Spoken.

5.3 Concluding remarks

On average, more registers in CORE have proportionally more densification than offline registers have. This is more evident in terms of phrasal postmodification, clausal postmodification and complementation. The involved registers in CORE exhibit their density strategies in clausal modifiers and noun complement clauses compared to expository writing in COCA. It therefore seems that popular online written registers are incorporating noun modification features on the clausal end of the spectrum.

Spoken was rarely distinguished from written registers in both COCA and CORE, except where obvious larger contrasts occurred in register types such as Academic. And, in all but clausal postmodification, Spoken did not have the lowest relative frequency of a modifying or complementising device where the expectation was that it should have.

While online registers are mostly distinguished from offline registers in terms of density, this often depends on the individual structures making up the modification and complementation types. Differences between outermost registers are also more pronounced than differences between registers in the middle rankings. Therefore, some online registers are not that much different from offline registers, even when they do have higher relative frequencies.

The same goes for frequency differences that evidently separate registers in terms of purpose. Many register types did show stark contrasts, but in some cases density did not distinguish many written registers (especially those in middle-rank positions). Further, where phrasal premodification and postmodification were concerned, registers generally seem to follow a cline from general to informational when ranked, but this was less evident in clausal postmodification and complementation.

5.4 Areas that need further exploration

Original texts for CORE are only available in their concordance format (unless searching for them using the links provided in metadata). It would be an interesting additional study to narrow the focus to one or two of the particular registers that show contrasting modification patterns across the groups of features and to compile an offline version of the texts that will still be available (the dreadful ‘page not found’ error) and to see whether different discourse styles are largely at play.

It would be interesting to further explore the interchange between *s*-genitives and *of*-possessives in the News registers.

5.5 Conclusion

This investigation has addressed the research question of whether and to what extent density characterises web-based registers compared to offline registers. While text purpose and various complex discourse factors influence noun phrase modification and complementation occurrences in registers, the online medium also affects language written on the web. The web's influence on the written form is partially brought to light by observing the frequency patterns of a select group of information-packaging features related to the noun phrase.

REFERENCES

Anderwald, L. 2007. ‘He rung the bell’ and ‘she drunk ale’ – non-standard past tense forms in traditional British dialects and on the internet. In: Hundt, M., Nesselhauf, N. & Biewer, C., eds. *Corpus linguistics and the web*. Amsterdam: Rodopi. pp. 271–285.

Atkinson, D. 2001. Scientific discourse across history: a combined multi-dimensional/rhetorical analysis of the *Philosophical Transactions of the Royal Society of London*. In: Conrad, S. & Biber, D., eds. *Variation in English: multi-dimensional studies.* New York: Routledge. pp. 45–65.

Baron, N.S. 2008. *Always on: language in an online and mobile world*. New York: Oxford University Press.

Beißwenger, M. 2008. Situated chat analysis as a window to the user’s perspective: aspects of temporal and sequential organization. *Language@Internet*, 5:1–19.

Biber, D. 1988. *Variation across speech and writing*. New York: Cambridge University Press.

Biber, D. 1995. *Dimensions of register variation: a cross-linguistic comparison.* Cambridge: Cambridge University Press.

Biber, D. 2001. On the complexity of discourse complexity. In: Conrad, S. & Biber, D., eds. *Variation in English: multi-dimensional studies*. New York: Routledge. pp. 215–240.

Biber, D. 2003. Compressed noun-phrase structure in newspaper discourse: the competing demands of popularization vs. economy. In: Aitchison, J. & Lewis, D.M., eds. *New media language*. London: Routledge. pp. 169–181.

Biber, D. 2006. *University language: a corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.

Biber, D. 2010. Corpus-based and corpus-driven analyses of language variation and use. In: Heine, B. & Narrog, H., eds. *The Oxford handbook of linguistic analysis*. London: Oxford University Press. pp. 159–191.

Biber, D. 2012. Register and discourse analysis. In: Gee, J.P. & Handford, M., eds. *The Routledge handbook of discourse analysis*. London: Routledge. pp. 191–208.

Biber, D. & Burges. 2001. Historical shifts in the language of women and men: gender differences in dramatic dialogue. In: Conrad, S. & Biber, D., eds. *Variation in English: multi-dimensional studies*. New York: Routledge. pp. 157–170.

Biber, D. & Clark, V. 2002. Historical shifts in modification patterns with complex noun phrase structures: how long can you go without a verb? In: Fanego, T., López-Couso, M.J. & Pérez-Guerra, J., eds. *English historical syntax and morphology*. Amsterdam: John Benjamins. pp. 43–66.

Biber, D. & Conrad, S. 2001. Introduction: multi-dimensional analysis and the study of register variation. In: Conrad, S. & Biber, D., eds. *Variation in English: multi-dimensional studies*. New York: Routledge. pp. 3–12.

Biber, D. & Conrad, S. 2003. Register variation: a corpus approach. In: Schiffrin, D., Tannen, D. & Hamilton, H.E., eds. *The handbook of discourse analysis*. Oxford: Blackwell. pp. 175–196.

Biber, D. & Conrad, S. 2009. *Register, genre, and style*. New York: Cambridge University Press.

Biber, D., Conrad, S. & Leech, G. 2002. *Longman student grammar of spoken and written English.* Harlow: Pearson Education.

Biber, D. & Egbert, J. 2016. Register variation on the searchable web: a multi-dimensional analysis. *Journal of English linguistics*, 44(2):95–137.

Biber, D., Egbert, J. & Davies, M. 2015. Exploring the composition of the searchable web: a corpus-based taxonomy of web registers. *Corpora*, 10(1):11–45.

Biber, D. & Finegan, E. 2001. Diachronic relations among speech-based and written registers in English. In: Conrad, S. & Biber, D., eds. *Variation in English: multi-dimensional studies*. New York: Routledge. pp. 66–83.

Biber, D. & Gray, B. 2011. Grammatical change in the noun phrase: the influence of written language use. *English language and linguistics*, 15(2):223–250.

Biber, D. & Gray, B. 2012. The competing demands of popularization vs. economy: written language in the age of mass literacy. In: Nevalainen, T. & Traugott, E.C., eds. *The Oxford handbook of the history of English*. New York: Oxford. pp. 314–328.

Biber, D. & Gray, B. 2016. *Grammatical complexity in academic English: linguistic change in writing*. Cambridge: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. 1999. *Longman grammar of spoken and written English*. Harlow: Pearson Education.

Brezina, V. 2018. *Statistics in corpus linguistics: a practical guide*. Cambridge: Cambridge University Press.

Crystal, D. 2006. *Language and the internet*. New York: Cambridge University Press.

Crystal, D. 2011. *Internet linguistics: a student guide*. Abingdon: Routledge.

Cumming, G. 2012. *Understanding the new statistics: effect sizes, confidence intervals, and meta-analysis*. New York: Routledge.

Cumming, G. 2014. The new statistics: why and how. *Psychological science*, 25(1):7–29. https://doi.org/10.1177/0956797613504966

Cutting, J. 2013. Spoken discourse. In: Hyland, K. & Paltridge, B., eds. *The Bloomsbury companion to discourse analysis*. London: Bloomsbury. pp. 155–170.

Davies, M. 2008-. *The Corpus of Contemporary American English (COCA): One billion words, 1990–2019*. Available online at https://www.english-corpora.org/coca/

Davies, J. 2013. Discourse and computer-mediated communication. In: Hyland, K. & Paltridge, B., eds. *The Bloomsbury companion to discourse analysis*. London: Bloomsbury. pp. 228–243.

Davies, M. 2013. *Corpus of Global Web-Based English: 1.9 billion words from speakers in 20 countries (GloWbE)*. Available online at https://www.english-corpora.org/glowbe/

Davies, M. 2016-. *Corpus of Online Registers of English (CORE)*. Available online at https://www.english-corpora.org/core/

Davies, M. & Fuchs, R. 2015. Expanding horizons in the study of World Englishes with the 1.9 billion word Global Web-based English Corpus (GloWbE). *English world-wide*, 36(1):1–28.

Egbert, J., Biber, D. & Davies, M. 2015. Developing a bottom-up, user-based method of web register classification. *Journal of the association for information science and technology*, 66(9):1817–1831.

Ellis, P.D. 2010. *The essential guide to effect sizes: statistical power, meta-analysis, and the interpretation of research results*. New York: Cambridge University Press.

Gabrielatos, C. & Marchi, A. 2012. *Keyness: appropriate metrics and practical issues.* Paper presented at Corpus-assisted Discourse Studies International Conference – University of Bologna, Italy. https://research.edgehill.ac.uk/en/publications/keyness-appropriate-metrics-and-practical-issues-2 Date of access: 1 Aug. 2020.

Gabrielatos, C. 2018. Keyness analysis: nature, metrics and techniques. In: Taylor, C. & Marchi, A., eds. *Corpus approaches to discourse: a critical review*. New York: Routledge. pp. 225–258.

Gray, B. 2013. More than discipline: uncovering multi-dimensional patterns of variation in academic research articles. *Corpora*, 8(2):153–181.

Gray, B. 2015. On the complexity of academic writing: disciplinary variation and structural complexity. In: Cortes, V. & Csomay, E., eds. *Corpus-based research in applied linguistics*. Amsterdam: John Benjamins. pp. 49–77.

Halliday, M.A.K. 1989. *Spoken and written language*. 2nd ed. Oxford: Oxford University Press.

Halsey, L.G. 2019. The reign of the p-value is over: what alternative analyses could we employ to fill the power vacuum? *Biology letters*, 15(5):1–8. https://doi.org/10.1098/rsbl.2019.0174

Herring, S.C. 2010. Computer-mediated conversation, part 1: introduction and overview. *Language@Internet,* 7:1–12. https://www.languageatinternet.org/articles/2010/2801 Date of access: 19 Mar. 2016.

Herring, S.C. 2013. Discourse in Web 2.0: familiar, reconfigured, and emergent. In: Tannen, D. & Trester, A.M., eds. *Discourse 2.0: language and new media*. Washington: Georgetown University Press. pp. 1–26.

Huddleston, R. 2002. Non-finite and verbless clauses. In: Huddleston, R. & Pullum, G.K., eds. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press. pp. 1171–1271.

Huddleston, R. & Pullum, G.K. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Huddleston, R., Pullum, G.K. & Peterson, P. 2002. Relative constructions and unbounded dependencies. In: Huddleston, R. & Pullum, G.K., eds. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press. pp. 1031–1096.

Johnson, E. 2015. Conversational writing: a multidimensional study of synchronous and supersynchronous computer-mediated communication. New York: Peter Lang. http://library.oapen.org/handle/20.500.12657/32731 Date of access: 9 Apr. 2016.

Kilgarriff, A. 2005. Language is never, ever, ever, random. *Corpus linguistics and linguistic theory*, 1(2):263–276. https://doi-org.nwulib.nwu.ac.za/10.1515/cllt.2005.1.2.263

Kilgarriff, A. 2009. Simple maths for keywords. In: Mahlberg, M., Díaz, V.G. & Smith, C., eds. *Proceedings of the Corpus Linguistics Conference,* CL2009, University of Liverpool, July. http://ucrel.lancs.ac.uk/publications/cl2009/

Kruger, H. & Van Rooy, B. 2016. Constrained language: a multidimensional analysis of translated English and a non-native indigenized variety of English. *English world-wide*, 37(1):26–57.

Leech, G., Hundt, M., Mair, C. & Smith, N. 2009. *Change in contemporary English: a grammatical study*. New York: Cambridge University Press.

Leech, G. & Smith, N. 2009. Change and constancy in linguistic change: how grammatical usage in written English evolved in the period 1931–1991 In: Renouf, A. & Kehoe, A., eds. *Corpus linguistics: refinements and reassessments*. Amsterdam: Rodopi: pp.173–200.

Lüdeling, A., Evert, S. & Baroni, M. 2007. Using web data for linguistic purposes. In: Hundt, M., Nesselhauf, N. & Biewer, C., eds. *Corpus linguistics and the web*. Amsterdam: Rodopi. pp. 7–24.

Mair, C. 2007. Change and variation in present-day English: integrating the analysis of closed corpora and web-based monitoring. In: Hundt, M., Nesselhauf, N. & Biewer, C., eds. *Corpus linguistics and the web*. Amsterdam: Rodopi. pp. 233–247.

McEnery, T. & Hardie, A. 2012. *Corpus linguistics: method, theory and practice*. Cambridge: Cambridge University Press.

McEnery, T. & Wilson, A. 2001. *Corpus linguistics: an introduction*. 2nd ed. Edinburgh: Edinburgh University Press.

Oakes, M.W. 1986. *Statistical inference: a commentary for the social and behavioural science*. Chichester: John Wiley & Sons.

Payne, J. & Huddleston, R. 2002. Nouns and noun phrases. In: Huddleston, R. & Pullum, G.K., eds. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press. pp. 323–523.

Pojanapunya, P. & Watson Todd, R. 2018. Log-likelihood and odds ratio: keyness statistics for different purposes of keyword analysis. *Corpus linguistics and linguistic theory*, 14(1):133–167.

Pullum, G.K. & Huddleston, R. 2002a. Adjectives and adverbs. In: Huddleston, R. & Pullum, G.K., eds. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press. pp. 525–595.

Pullum, G.K. & Huddleston, R. 2002b. Prepositions and prepositional phrases. In: Huddleston, R. & Pullum, G.K., eds. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press. pp. 597–661.

Reppen, R. 2001. Register variation in student and adult speech and writing. In: Conrad, S. & Biber, D., eds. *Variation in English: multi-dimensional studies*. New York: Routledge. pp. 187–199.

Rey, J.M. 2001. Changing gender roles in popular culture: dialogue in Star Trek episodes from 1966 to 1993. In: Conrad, S. & Biber, D., eds. *Variation in English: multi-dimensional studies*. New York: Routledge. pp. 138–156.

Rohdenburg, G. 2007. Determinants of grammatical variation in English and the formation/confirmation of linguistic hypotheses by means of internet data. In: Hundt, M., Nesselhauf, N. & Biewer, C., eds. *Corpus linguistics and the web*. Amsterdam: Rodopi. pp. 191–209.

Rosenbach, A. 2007. Exploring constructions on the web: a case study. In: Hundt, M., Nesselhauf, N. & Biewer, C., eds. *Corpus linguistics and the web*. Amsterdam: Rodopi. pp. 167–190.

Seargeant, P. & Tagg, C. 2011. English on the internet and a ‘post-varieties’ approach to language. *World Englishes*, 30(4):496–514.

Squires, L. 2010. Enregistering internet language. *Language in Society*, 39(4):457–492.

Tagliamonte, S.A. 2016. So sick or so cool? The language of youth on the internet. *Language in society*, 45:1–32.

Tannen, D. 2013. The medium is the metamessage: conversational style in new media interaction. In: Tannen, D. & Trester, A.M., eds. *Discourse 2.0: language and new media*. Washington: Georgetown University Press. pp. 99–118.

Taylor, C. 2018. Similarity. In: Taylor, C. & Marchi, A., eds. *Corpus approaches to discourse: a critical review*. Routledge: New York. pp.19–37.

Titak, A. & Roberson, A. 2013. Dimensions of web registers: an exploratory multi-dimensional comparison. *Corpora*, 8(2):235–260. DOI: 10.3366/cor.2013.0042

Van Rooy, B. & Terblanche, L. 2009. A multi-dimensional analysis of a learner corpus. In: Renouf, A. & Kehoe, A., eds. *Corpus linguistics: refinements and reassessments*. Rodopi: Amsterdam. pp. 239–254.

West, L. & Trester, A.M. 2013. Facework on Facebook: conversations on social media. In: Tannen, D. & Trester, A.M., eds. *Discourse 2.0: language and new media*. Washington: Georgetown University Press. pp. 133–154.

APPENDIX

The %DIFF score for each register pair is ranked from largest difference to smallest difference. RRF value is included for the purpose of providing an interpretable difference in terms of size e.g. 1.50 = one and a half times more. These two values are equal in the size of the difference. The register in the left position of each pair contain the higher relative frequency for the relevant construction.

Table A.1 Pairwise comparison of common nouns as prenominal modifier

FAQ_Historical Article	123,66	2,24
FAQ_Fiction	121,05	2,21
FAQ_Spoken	104,99	2,05
Academic_Historical Article	96,07	1,96
Academic_Fiction	93,78	1,94
Academic_Spoken	79,70	1,80
FAQ_Personal Blog	74,66	1,75
CORE News_Historical Article	63,91	1,64
CORE News_Fiction	62,00	1,62
FAQ_Interactive Discussion	61,81	1,62
News_Historical Article	60,78	1,61
News_Fiction	58,91	1,59
Academic_Personal Blog	53,11	1,53
Magazine_Historical Article	51,05	1,51
CORE News_Spoken	50,23	1,50
Magazine_Fiction	49,29	1,49
FAQ_Magazine	48,07	1,48
News_Spoken	47,36	1,47
Academic_Interactive Discussion	41,85	1,42

FAQ_News	39,11	1,39
Magazine_Spoken	38,44	1,38
Interactive Discussion_Historical Article	38,22	1,38
Interactive Discussion_Fiction	36,61	1,37
FAQ_CORE News	36,46	1,36
Academic_Magazine	29,80	1,30
Personal Blog_Historical Article	28,06	1,28
CORE News_Personal Blog	28,00	1,28
Interactive Discussion_Spoken	26,69	1,27
Personal Blog_Fiction	26,56	1,27
News_Personal Blog	25,55	1,26
Academic_News	21,95	1,22
Academic_CORE News	19,62	1,20
CORE News_Interactive Discussion	18,58	1,19
Magazine_Personal Blog	17,96	1,18
Personal Blog_Spoken	17,37	1,17
News_Interactive Discussion	16,32	1,16
FAQ_Academic	14,07	1,14
Magazine_Interactive Discussion	9,28	1,09
Spoken_Historical Article	9,11	1,09
CORE News_Magazine	8,51	1,09
Interactive Discussion_Personal Blog	7,94	1,08
Spoken_Fiction	7,83	1,08
News_Magazine	6,44	1,06
CORE News_News	1,94	1,02
Fiction_Historical Article	1,18	1,01

Table A.2 Pairwise comparison of proper nouns as prenominal modifier

CORE News_Fiction	447,77	5,48
Historical Article_Fiction	375,67	4,76
News_Fiction	324,85	4,25
Spoken_Fiction	236,50	3,37
CORE News_Academic	199,30	2,99
CORE News_Interactive Discussion	198,70	2,99
FAQ_Fiction	179,41	2,79
CORE News_Personal Blog	167,91	2,68
Historical Article_Academic	159,91	2,60
Historical Article_Interactive Discussion	159,38	2,59
CORE News_Magazine	137,39	2,37
Historical Article_Personal Blog	132,65	2,33
News_Academic	132,14	2,32
News_Interactive Discussion	131,67	2,32
Magazine_Fiction	130,75	2,31
News_Personal Blog	107,79	2,08
Historical Article_Magazine	106,14	2,06
Personal Blog_Fiction	104,46	2,04
CORE News_FAQ	96,05	1,96
News_Magazine	84,12	1,84
Spoken_Academic	83,86	1,84
Spoken_Interactive Discussion	83,49	1,83
Interactive Discussion_Fiction	83,39	1,83
Academic_Fiction	83,02	1,83

Historical Article_FAQ	70,24	1,70
Spoken_Personal Blog	64,58	1,65
CORE News_Spoken	62,79	1,63
FAQ_Academic	52,67	1,53
FAQ_Interactive Discussion	52,36	1,52
News_FAQ	52,05	1,52
Spoken_Magazine	45,83	1,46
Historical Article_Spoken	41,36	1,41
FAQ_Personal Blog	36,66	1,37
CORE News_News	28,93	1,29
Spoken_News	26,25	1,26
Magazine_Academic	26,08	1,26
Magazine_Interactive Discussion	25,83	1,26
FAQ_Magazine	21,09	1,21
Spoken_FAQ	20,43	1,20
CORE News_Historical Article	15,16	1,15
Magazine_Personal Blog	12,86	1,13
Historical Article_News	11,96	1,12
Personal Blog_Academic	11,72	1,12
Personal Blog_Interactive Discussion	11,49	1,11
Interactive Discussion_Academic	0,20	1,00

Table A.3 Pairwise comparison of adjectives as prenominal modifier

Academic_Interactive Discussion	50,02	1,50
Academic_Spoken	48,53	1,49
Academic_Fiction	46,66	1,47

Academic_Personal Blog	36,04	1,36
Academic_News	28,78	1,29
Historical Article_Interactive Discussion	28,02	1,28
Historical Article_Spoken	26,75	1,27
Magazine_Interactive Discussion	25,99	1,26
Academic_FAQ	25,77	1,26
Historical Article_Fiction	25,16	1,25
Magazine_Spoken	24,74	1,25
CORE News_Interactive Discussion	23,90	1,24
Magazine_Fiction	23,17	1,23
CORE News_Spoken	22,67	1,23
CORE News_Fiction	21,13	1,21
Academic_CORE News	21,08	1,21
FAQ_Interactive Discussion	19,28	1,19
Academic_Magazine	19,07	1,19
FAQ_Spoken	18,10	1,18
Academic_Historical Article	17,18	1,17
FAQ_Fiction	16,61	1,17
News_Interactive Discussion	16,49	1,16
Historical Article_Personal Blog	16,09	1,16
News_Spoken	15,34	1,15
Magazine_Personal Blog	14,24	1,14
News_Fiction	13,88	1,14
CORE News_Personal Blog	12,35	1,12
Personal Blog_Interactive Discussion	10,28	1,10
Historical Article_News	9,90	1,10

Personal Blog_Spoken	9,18	1,09
FAQ_Personal Blog	8,16	1,08
Magazine_News	8,15	1,08
Personal Blog_Fiction	7,81	1,08
Historical Article_FAQ	7,33	1,07
CORE News_News	6,36	1,06
News_Personal Blog	5,63	1,06
Magazine_FAQ	5,62	1,06
CORE News_FAQ	3,87	1,04
Historical Article_CORE News	3,33	1,03
FAQ_News	2,40	1,02
Fiction_Interactive Discussion	2,29	1,02
Magazine_CORE News	1,68	1,02
Historical Article_Magazine	1,61	1,02
Fiction_Spoken	1,27	1,01
Spoken_Interactive Discussion	1,00	1,01

Table A.4 Pairwise comparison of *of*-phrases as postnominal modifier

Historical Article_News	100,48	2,00
Historical Article_Interactive Discussion	82,28	1,82
Historical Article_Spoken	76,19	1,76
Academic_News	70,87	1,71
Historical Article_Fiction	65,37	1,65
Historical Article_Magazine	58,66	1,59
Academic_Interactive Discussion	55,36	1,55
Historical Article_Personal Blog	51,02	1,51

Academic_Spoken	50,17	1,50
FAQ_News	49,45	1,49
Historical Article_CORE News	42,96	1,43
Academic_Fiction	40,95	1,41
CORE News_News	40,23	1,40
FAQ_Interactive Discussion	35,88	1,36
Academic_Magazine	35,23	1,35
Historical Article_FAQ	34,14	1,34
Personal Blog_News	32,74	1,33
FAQ_Spoken	31,34	1,31
Academic_Personal Blog	28,72	1,29
CORE News_Interactive Discussion	27,50	1,27
Magazine_News	26,36	1,26
FAQ_Fiction	23,28	1,23
CORE News_Spoken	23,24	1,23
Academic_CORE News	21,85	1,22
Fiction_News	21,23	1,21
Personal Blog_Interactive Discussion	20,69	1,21
FAQ_Magazine	18,28	1,18
Historical Article_Academic	17,32	1,17
Personal Blog_Spoken	16,66	1,17
CORE News_Fiction	15,67	1,16
Magazine_Interactive Discussion	14,89	1,15
Academic_FAQ	14,33	1,14
Spoken_News	13,79	1,14
FAQ_Personal Blog	12,59	1,13

Magazine_Spoken	11,05	1,11
CORE News_Magazine	10,98	1,11
Fiction_Interactive Discussion	10,22	1,10
Interactive Discussion_News	9,98	1,10
Personal Blog_Fiction	9,50	1,10
FAQ_CORE News	6,58	1,07
Fiction_Spoken	6,54	1,07
CORE News_Personal Blog	5,64	1,06
Personal Blog_Magazine	5,05	1,05
Magazine_Fiction	4,23	1,04
Spoken_Interactive Discussion	3,46	1,03

Table A.5 Pairwise comparison of *in*-phrase as postnominal modifier

Academic_Interactive Discussion	136,25	2,36
Historical Article_Interactive Discussion	121,45	2,21
Academic_Personal Blog	115,58	2,16
Historical Article_Personal Blog	102,08	2,02
CORE News_Interactive Discussion	92,22	1,92
Academic_Fiction	91,84	1,92
Historical Article_Fiction	79,82	1,80
CORE News_Personal Blog	75,40	1,75
Academic_Spoken	66,40	1,66
News_Interactive Discussion	63,21	1,63
Magazine_Interactive Discussion	58,07	1,58
CORE News_Fiction	56,09	1,56
Historical Article_Spoken	55,98	1,56

Academic_FAQ	55,55	1,56
FAQ_Interactive Discussion	51,87	1,52
Academic_Magazine	49,46	1,49
News_Personal Blog	48,93	1,49
Historical Article_FAQ	45,81	1,46
Academic_News	44,75	1,45
Magazine_Personal Blog	44,24	1,44
Spoken_Interactive Discussion	41,97	1,42
Historical Article_Magazine	40,10	1,40
FAQ_Personal Blog	38,59	1,39
Historical Article_News	35,68	1,36
CORE News_Spoken	35,39	1,35
News_Fiction	32,53	1,33
Spoken_Personal Blog	29,55	1,30
Magazine_Fiction	28,35	1,28
CORE News_FAQ	26,57	1,27
FAQ_Fiction	23,32	1,23
Fiction_Interactive Discussion	23,15	1,23
Academic_CORE News	22,90	1,23
CORE News_Magazine	21,61	1,22
CORE News_News	17,77	1,18
Spoken_Fiction	15,28	1,15
Historical Article_CORE News	15,21	1,15
News_Spoken	14,96	1,15
Fiction_Personal Blog	12,37	1,12
Magazine_Spoken	11,34	1,11

Personal Blog_Interactive Discussion	9,59	1,10
News_FAQ	7,47	1,07
FAQ_Spoken	6,97	1,07
Academic_Historical Article	6,68	1,07
Magazine_FAQ	4,08	1,04
News_Magazine	3,26	1,03

Table A.6 Pairwise comparison of *for*-phrase as postnominal modifier

Academic_Fiction	218,55	3,19
FAQ_Fiction	209,56	3,10
CORE News_Fiction	180,79	2,81
Historical Article_Fiction	150,73	2,51
Personal Blog_Fiction	133,36	2,33
News_Fiction	124,39	2,24
Spoken_Fiction	101,62	2,02
Interactive Discussion_Fiction	96,38	1,96
Magazine_Fiction	83,94	1,84
Academic_Magazine	73,18	1,73
FAQ_Magazine	68,29	1,68
Academic_Interactive Discussion	62,21	1,62
Academic_Spoken	58,00	1,58
FAQ_Interactive Discussion	57,64	1,58
FAQ_Spoken	53,54	1,54
CORE News_Magazine	52,65	1,53
CORE News_Interactive Discussion	42,99	1,43
Academic_News	41,96	1,42

CORE News_Spoken	39,27	1,39
FAQ_News	37,96	1,38
Academic_Personal Blog	36,51	1,37
Historical Article_Magazine	36,31	1,36
FAQ_Personal Blog	32,66	1,33
Historical Article_Interactive Discussion	27,68	1,28
Academic_Historical Article	27,05	1,27
Personal Blog_Magazine	26,86	1,27
CORE News_News	25,14	1,25
Historical Article_Spoken	24,36	1,24
FAQ_Historical Article	23,46	1,23
News_Magazine	21,99	1,22
CORE News_Personal Blog	20,33	1,20
Personal Blog_Interactive Discussion	18,83	1,19
Personal Blog_Spoken	15,74	1,16
News_Interactive Discussion	14,27	1,14
Academic_CORE News	13,45	1,13
CORE News_Historical Article	11,99	1,12
Historical Article_News	11,74	1,12
News_Spoken	11,30	1,11
FAQ_CORE News	10,25	1,10
Spoken_Magazine	9,61	1,10
Historical Article_Personal Blog	7,44	1,07
Interactive Discussion_Magazine	6,76	1,07
Personal Blog_News	4,00	1,04
Academic_FAQ	2,90	1,03

Spoken_Interactive Discussion	2,67	1,03

Table A.7 Pairwise comparison of *that*-relative clause as postnominal modifier

Spoken_Historical Article	228,49	3,28
Personal Blog_Historical Article	126,91	2,27
Spoken_Fiction	105,53	2,06
Spoken_CORE News	104,17	2,04
Magazine_Historical Article	98,94	1,99
Interactive Discussion_Historical Article	98,73	1,99
Spoken_Academic	93,97	1,94
News_Historical Article	86,92	1,87
FAQ_Historical Article	86,91	1,87
Spoken_FAQ	75,75	1,76
Spoken_News	75,74	1,76
Academic_Historical Article	69,35	1,69
Spoken_Interactive Discussion	65,29	1,65
Spoken_Magazine	65,12	1,65
CORE News_Historical Article	60,89	1,61
Fiction_Historical Article	59,82	1,60
Spoken_Personal Blog	44,76	1,45
Personal Blog_Fiction	41,98	1,42
Personal Blog_CORE News	41,04	1,41
Personal Blog_Academic	33,99	1,34
Magazine_Fiction	24,48	1,24
Interactive Discussion_Fiction	24,35	1,24
Magazine_CORE News	23,65	1,24

Interactive Discussion_CORE News	23,52	1,24
Personal Blog_FAQ	21,40	1,21
Personal Blog_News	21,40	1,21
Magazine_Academic	17,47	1,17
Interactive Discussion_Academic	17,35	1,17
News_Fiction	16,95	1,17
FAQ_Fiction	16,95	1,17
News_CORE News	16,18	1,16
FAQ_CORE News	16,17	1,16
Personal Blog_Interactive Discussion	14,18	1,14
Personal Blog_Magazine	14,06	1,14
News_Academic	10,37	1,10
FAQ_Academic	10,37	1,10
Magazine_FAQ	6,44	1,06
Magazine_News	6,43	1,06
Interactive Discussion_FAQ	6,33	1,06
Interactive Discussion_News	6,32	1,06
Academic_Fiction	5,96	1,06
Academic_CORE News	5,26	1,05
CORE News_Fiction	0,67	1,01
Magazine_Interactive Discussion	0,10	1,00
News_FAQ	0,00	1,00

Table A.8 Pairwise comparison of *which*-relative clause as postnominal modifier

Historical Article_News	3085,67	31,86
Historical Article_Magazine	2271,19	23,71

FAQ_News	2211,40	23,11
Interactive Discussion_News	1766,29	18,66
FAQ_Magazine	1620,45	17,20
Personal Blog_News	1469,88	15,70
Historical Article_Fiction	1298,46	13,98
Interactive Discussion_Magazine	1289,13	13,89
CORE News_News	1284,83	13,85
Personal Blog_Magazine	1068,51	11,69
CORE News_Magazine	930,77	10,31
FAQ_Fiction	914,67	10,15
Historical Article_Academic	720,44	8,20
Interactive Discussion_Fiction	719,27	8,19
Personal Blog_Fiction	589,16	6,89
CORE News_Fiction	507,92	6,08
FAQ_Academic	495,28	5,95
Spoken_News	479,15	5,79
Historical Article_Spoken	450,06	5,50
Interactive Discussion_Academic	380,65	4,81
Spoken_Magazine	331,08	4,31
Personal Blog_Academic	304,31	4,04
FAQ_Spoken	299,10	3,99
Academic_News	288,29	3,88
CORE News_Academic	256,65	3,57
Interactive Discussion_Spoken	222,24	3,22
Academic_Magazine	189,01	2,89
Personal Blog_Spoken	171,06	2,71

Spoken_Fiction	154,24	2,54
CORE News_Spoken	139,11	2,39
Historical Article_CORE News	130,04	2,30
Fiction_News	127,80	2,28
Historical Article_Personal Blog	102,92	2,03
Historical Article_Interactive Discussion	70,70	1,71
Academic_Fiction	70,45	1,70
Fiction_Magazine	69,56	1,70
FAQ_CORE News	66,91	1,67
Spoken_Academic	49,16	1,49
FAQ_Personal Blog	47,23	1,47
Historical Article_FAQ	37,82	1,38
Interactive Discussion_CORE News	34,77	1,35
Magazine_News	34,35	1,34
FAQ_Interactive Discussion	23,85	1,24
Interactive Discussion_Personal Blog	18,88	1,19
Personal Blog_CORE News	13,36	1,13

Table A.9 Pairwise comparison of *to*-clause as postnominal modifier

Personal Blog_Historical Article	197,41	2,97
Personal Blog_Academic	173,47	2,73
Interactive Discussion_Historical Article	162,03	2,62
Interactive Discussion_Academic	140,94	2,41
Personal Blog_News	112,90	2,13
FAQ_Historical Article	111,73	2,12
FAQ_Academic	94,69	1,95

Personal Blog_CORE News	92,66	1,93
Interactive Discussion_News	87,58	1,88
Personal Blog_Fiction	84,56	1,85
Magazine_Historical Article	79,20	1,79
Personal Blog_Spoken	75,18	1,75
Spoken_Historical Article	69,77	1,70
Interactive Discussion_CORE News	69,74	1,70
Personal Blog_Magazine	65,97	1,66
Magazine_Academic	64,77	1,65
Interactive Discussion_Fiction	62,61	1,63
Fiction_Historical Article	61,14	1,61
Spoken_Academic	56,11	1,56
CORE News_Historical Article	54,37	1,54
Interactive Discussion_Spoken	54,35	1,54
FAQ_News	51,57	1,52
Fiction_Academic	48,17	1,48
Interactive Discussion_Magazine	46,23	1,46
CORE News_Academic	41,95	1,42
Personal Blog_FAQ	40,46	1,40
News_Historical Article	39,70	1,40
FAQ_CORE News	37,16	1,37
FAQ_Fiction	31,39	1,31
News_Academic	28,45	1,28
Magazine_News	28,28	1,28
FAQ_Spoken	24,72	1,25
Interactive Discussion_FAQ	23,76	1,24

Spoken_News	21,53	1,22
FAQ_Magazine	18,16	1,18
Magazine_CORE News	16,08	1,16
Fiction_News	15,35	1,15
Personal Blog_Interactive Discussion	13,50	1,13
Magazine_Fiction	11,20	1,11
CORE News_News	10,51	1,11
Spoken_CORE News	9,98	1,10
Academic_Historical Article	8,75	1,09
Magazine_Spoken	5,55	1,06
Spoken_Fiction	5,35	1,05
Fiction_CORE News	4,39	1,04

Table A.10 Pairwise comparison of *ed*/*en*-clause as postnominal modifier

FAQ_Spoken	291,53	3,92
FAQ_Fiction	225,18	3,25
Academic_Spoken	204,57	3,05
FAQ_Interactive Discussion	189,25	2,89
Historical Article_Spoken	180,56	2,81
CORE News_Spoken	163,06	2,63
FAQ_Personal Blog	157,51	2,58
Academic_Fiction	152,96	2,53
Historical Article_Fiction	133,01	2,33
Academic_Interactive Discussion	125,01	2,25
FAQ_News	122,60	2,23
CORE News_Fiction	118,48	2,18

FAQ_Magazine	110,33	2,10
Historical Article_Interactive Discussion	107,27	2,07
Academic_Personal Blog	100,32	2,00
CORE News_Interactive Discussion	94,34	1,94
Magazine_Spoken	86,14	1,86
Historical Article_Personal Blog	84,53	1,85
News_Spoken	75,89	1,76
Academic_News	73,16	1,73
CORE News_Personal Blog	73,02	1,73
Academic_Magazine	63,62	1,64
Historical Article_News	59,51	1,60
Magazine_Fiction	54,60	1,55
Personal Blog_Spoken	52,04	1,52
Historical Article_Magazine	50,72	1,51
CORE News_News	49,56	1,50
FAQ_CORE News	48,83	1,49
News_Fiction	46,08	1,46
CORE News_Magazine	41,32	1,41
FAQ_Historical Article	39,55	1,40
Magazine_Interactive Discussion	37,52	1,38
Interactive Discussion_Spoken	35,36	1,35
News_Interactive Discussion	29,94	1,30
FAQ_Academic	28,55	1,29
Personal Blog_Fiction	26,28	1,26
Magazine_Personal Blog	22,43	1,22
Fiction_Spoken	20,40	1,20

Academic_CORE News	15,78	1,16
News_Personal Blog	15,68	1,16
Interactive Discussion_Fiction	12,42	1,12
Personal Blog_Interactive Discussion	12,33	1,12
Academic_Historical Article	8,56	1,09
Historical Article_CORE News	6,65	1,07
Magazine_News	5,83	1,06

Table A.11 Pairwise comparison of *that*-clause as postnominal complement

CORE News_Spoken	214,15	3,14
CORE News_Spoken	188,54	2,89
Personal Blog_Spoken	147,80	2,48
Spoken_Fiction	144,88	2,45
CORE News_Spoken	134,09	2,34
Personal Blog_Spoken	127,60	2,28
Interactive Discussion_Spoken	125,66	2,26
Spoken_News	124,91	2,25
Interactive Discussion_Spoken	107,26	2,07
Historical Article_Spoken	107,06	2,07
Academic_Fiction	101,09	2,01
Historical Article_Spoken	90,18	1,90
FAQ_Spoken	85,20	1,85
Academic_News	84,70	1,85
Personal Blog_Spoken	84,65	1,85
Spoken_Magazine	82,47	1,82
FAQ_Spoken	70,10	1,70

CORE News_FAQ	69,63	1,70
Interactive Discussion_Spoken	68,15	1,68
CORE News_Spoken	56,22	1,56
Historical Article_Spoken	54,29	1,54
CORE News_Spoken	51,72	1,52
Academic_Magazine	49,84	1,50
CORE News_Interactive Discussion	39,22	1,39
FAQ_Spoken	38,00	1,38
Magazine_Fiction	34,20	1,34
Personal Blog_Spoken	33,80	1,34
Spoken_FAQ	32,22	1,32
CORE News_Spoken	28,29	1,28
CORE News_Personal Blog	26,78	1,27
Magazine_Spoken	23,26	1,23
Personal Blog_Spoken	23,23	1,23
Interactive Discussion_Spoken	21,85	1,22
Spoken_Academic	21,77	1,22
Personal Blog_Spoken	19,68	1,20
Spoken_Historical Article	18,26	1,18
Interactive Discussion_Spoken	12,22	1,12
Historical Article_FAQ	11,81	1,12
Personal Blog_Interactive Discussion	9,81	1,10
Interactive Discussion_Spoken	8,98	1,09
News_Fiction	8,88	1,09
Academic_FAQ	8,58	1,09
Spoken_Interactive Discussion	8,52	1,09

Historical Article_Spoken	2,97	1,03
Personal Blog_Spoken	1,19	1,01