THE SEMANTIC VARIATION OF THE SEMI-MODAL *NEED TO* IN CONTEMPORARY AMERICAN ENGLISH - A CORPUS-BASED STUDY

Niko Aaltonen, 251017
Master's Thesis
School of Humanities
English Language and Culture
University of Eastern Finland
Fall 2020

ITÄ-SUOMEN YLIOPISTO – UNIVERSITY OF EASTERN FINLAND

Faculty	**School**
Philosophical Faculty	School of Humanities

Author
Niko Aaltonen

Title
The Semantic Variation of the Semi-modal *Need to* in Contemporary American English – A Corpus-based Study

Main subject	**Level**	**Date**	**Number of pages**
English language and culture	Master's thesis	1.12.2020	82

Abstract

This thesis examines the semi-modal *need to* in contemporary American English using corpus-based methods. *Need to* is a semi-modal auxiliaxy verb that has gained popularity in the field of modal auxiliaries in the recent years, and is thus an interesting topic to study. *Need to* is a modal verb that is linked to the semantic notions of necessity and obligation, and may communicate these functions with different level's of the speaker's involvement. Moreover, this thesis takes into account the concept of modality, aiming to provide a concise definition for both the main term and its subcategories, dynamic, deontic, and epistemic modality. *Need to* is examined within the framework of these linguistic concepts.

The analysis in practice is conducted with the methodology provided by two earlier studies, Coates (1983) and Nokkonen (2015). Coates, in her study, examines the core modal *must*, which functions in similary ways to *need to*. She presents a fuzzy set theory, which argues that modality is a gradient phenomenon. This theory is applied to need to in this study in a similar manner with Nokkonen (2015). Nokkonen, moreover, provides a semantic categorization which is linked to Coates' ideas. This categorization is also used in this thesis when analyzing the instances of *need to*.

This study uses the 2006 corpus of American English (AmE06) as source. The AmE06 is a corpus compiled at the University of Lancaster by Amanda Potts and Paul Baker it is compiled in a similar manner to the Brown family of corpora, and includes roughly a million words of written American English mostly from 2005–2007. All the relevant instances found from the corpus are analyzed manually in their immediate context using a set of parameters developed by Coates (1983).

There are 248 instances of the semi-modal *need to* in the AmE06 corpus. Most instances are found in the fiction and general prose sections of the corpus. The analysis is divided into subsections base on the subject type. Most common subject type is the third person animate subject, although all subject types are found in the data. The instances are mostly weak, and situate themselves on the skirt in Coates' fuzzy set. There are some strong instances in the core or near it. Semantically, *need to* in the AmE06 covers a wide range of different functions, from strong, direct obligations to simply stating the objective facts or properties of an inanimate subject. There were no epistemic instances of *need to* found in the data.

ITÄ-SUOMEN YLIOPISTO – UNIVERSITY OF EASTERN FINLAND

Tiedekunta	**Osasto**
Filosofinen tiedekunta	Hunanistinen osasto

Tekijä
Niko Aaltonen

Työn nimi
The Semantic Variation of the Semi-modal *Need to* in Contemporary American English – A Corpus-based Study

Pääaine	**Työn laji**	**Päivämäärä**	**Sivumäärä**
Englannin kieli ja kulttuuri	Pro gradu -tutkielma	1.12.2020	82

Tiivistelmä

Tämän tutkielman aiheena on tutkia kuinka puolimodaaliapuverbi *need to* käyttäytyy amerikanenglannissa, käyttäen apuna korpustutkimuksen metodeja. Aikaisempien tutkimuksien mukaan *need to* on viime vuosina kasvattanut suosiotaan modaaliapuverbien kentässä ja on siksi mielenkiintoinen tutkimuksen kohde. *Need to* on ilmaisu, joka on kytköksissä semanttisiin käsitteisiin välttämättömyys ja velvollisuus, ja sitä puhuja voi ilmaista näitä erilaisella sitoutumisen tai osallistumisen voimakkuudella. Tämä tutkimus ottaa myös huomioon modaalisuuden käsitteen, pyrkii määrittelemään sen, sekä sen alakäsitteet dynaaminen, deonttinen ja episteeminen modaalisuus, mahdollisimman tarkasti. *Need to* on tarkastelun kohteena näiden käsitteiden viitekehyksessä.

Käytännön analyysi pohjautuu erityisesti kahteen aiempaan aiheesta tehtyyn tutkimukseen, Coates (1983) ja Nokkonen (2015). Coates keskittyy omassa tutkimuksessaan englannin kielen modaaliapuverbiin must, joka on verrattavissa puolimodaaliin need to. Hän esittelee ns. fuzzy set -teorian, jonka perusteella modaalisuuden käsite on jatkumo, eivätkä alakäsitteiden rajat ole ehdottomat. Tätä teoriaa hyödynnetään tämän tutkimuksen analyysiosuudessa samaan tapaan kuten Nokkonen (2015) on tehnyt. Nokkonen on myös omassa tutkimuksessaan luonut semanttisen kategorisoinnin Coatesin teorian pohjalta, ja tätä kategorisointia käytetään myös tässä tutkimuksessa.

Tutkimuksen aineistona käytetään vuoden 2006 amerikanenglannin korpusta (AmE06), jonka ovat koonneet Amanda Potts ja Paul Baker Lancasterin yliopistosta. Korpus on koostettu yhtenevällä tavalla Brown-korpusperheen kanssa, ja siinä on noin miljoona sanaa kirjoitettua Amerikanenglantia vuoden 2006 ympäriltä. Kaikki esiintymät korpuksesta analysoitiin manuaalisesti yksitellen niiden välittömässä kontekstissa, käyttäen apuna parametrejä, joita Coates (1983) hyödyntää omassa tutkimuksessaan.

Puolimodaalia *need to* löytyy aineistosta yhteensä 248 kertaa. Suurin osa osumista löytyy fiktio- ja yleinen proosa -osioista. Osumien analyysi on jaettu alalukuihin subjektityyppien perusteella. Yleisin subjektityyppi on kolmannen persoonan elollinen subjekti. Suurin osa tapauksista edustaa heikkoa deonttista modaalisuutta, ja sijoittuu Coatesin jatkumon reunalle. Aineistosta löytyy myös muutamia tapauksia, jotka edustavat voimakasta deonttisuutta, ja osuvat jatkumon keskelle tai sen läheisyyteen. Semanttisesta näkökulmasta *need to* kattaa leveän skaalan erilaisia funktioita, suorista käskyistä elottomien esineiden ominaisuuksien kuvaamiseen. Aineistosta ei löytynyt tapauksia, joissa *need to* olisi esiintynyt episteemisessä merkityksessä.

Contents

1. Introduction

The aim of this paper is to investigate the semantic variation of the semi-modal auxiliary verb *need to* in contemporary American English using corpus-based methods. Modal auxiliary verbs, such as *must*, *would*, *ought*, and *should* are a group of auxiliary verbs that are used to talk about situations that are not factual or not actualized (Aarts 2011: 275–276). They are related to the semantic concepts such as *probability*, *necessity*, or *obligation*. For example:

        (1)   She **must** be there by 9 am.

It is often assumed by non-professional linguists that grammatical change in a language is slow or even non-existent (Leech 2003: 223). This may be partly explained by the analytical nature of English, as little of the grammar is audible or visible directly (Leech et al. 2009: 7). Until fairly recently the study of grammatical change relied on unsystematic and impressionistic sources for information (Collins 2015: 1). However, since the introduction of electronic corpora, analyzing large amounts of text from varying time-spans and literary genres has become possible, and thus slowly progressing grammatical changes may be observed and studied.

Modal auxiliary verbs are a closed group, and they share properties that distinguish them from the class of open lexical verbs. However, there are verbs that can be dually categorized, i.e., they share some of the properties of the modal verbs, but also function as lexical verbs, such as the verb *have* in the following two examples (Huddleston and Pullum 2002: 92). Such expressions form the group of *emerging modals*, or *semi-modals*, which is the term applied in this paper.

| Auxiliary |
(2)   They **had** finished.

| Lexical |
(3)   They **had** a fight.

Several linguists (see, e.g., Leech et al. 2009) have recently noted that the English *modal system*, the grammatical notion which refers to the use of modal auxiliaries, is showing signs of change. There is a relevantly high consensus among scholars that the traditional English core modal auxiliaries (*can, could, may, must, need, ought, shall, should, will,* and *would*) are on the decline, while new modal expressions are emerging, e.g., the semi-modals such as *have to*, *be going to*, and *need to* (Leech et al. 2009: 99). Similar results have been presented by Leech (2003), Johansson (2013), and Daugs (2017) who all have observed changes in the modal system via corpus-based methods. Millar (2009: 191) states that while the English core modals are decreasing in frequency, the overall pattern is one of growth. Nokkonen (2015: 75) comments that since the reasons behind this phenomenon remain unclarified, the domain of modal auxiliaries seems to be a promising research area.

Especially interesting is the case of one expression which is currently emerging, namely the semi-modal *need to*. While, e.g., Millar (2003: 196), Leech (2003: 230), Johansson (2003: 374), and Smith (2003: 249) acknowledge in their respective studies that the frequency of *need to* has increased significantly in both British and American English in the past decades, there has been a lack of more in-depth analysis of *need to* and its semantics until Nokkonen (2015). Nokkonen, in her dissertation, has concentrated specifically on the semi-modal *need to* and its different semantical and pragmatical functions in various British English corpora with a mainly sociolinguistic approach. She found out that the semantic and pragmatic variation of *need to* depends primarily on the subject, both type and person, in addition with the authority structure

in the speech situation (Nokkonen 2015: 235). Her study is has been influential to the progress of writing this thesis by providing a theoretical offset and comparable results.

There is a gap in previous research as there is no previous comprehensive semantic analysis on *need to* in an American context. Hence, this thesis aims to describe the semantic variation of *need to* in contemporary American English with corpus-based methods, utilizing the semantic profile, i.e., categorization of the different uses of *need to* according to its semantics, created by Nokkonen (2015: 121–123, 153–156). Another framework, developed for the analysis of the core modal *must* by Coates (1983), will be utilized in the analysis. This framework was also used by Nokkonen (2015) as a basis for her analysis, and provides tools for placing instances of *need to* on a strong-weak continuum in accordance with the strength of the sense of obligation in the utterance. The material for this study will consist of corpus data from the 2006 corpus of American English (AmE06), compiled at the University of Lancaster, which is freely accessible online, see https://cqpweb.lancs.ac.uk/ame06/. The research question of this thesis is as follows:

- How does the semi-modal **need to** vary semantically in the 2006 corpus of American English.

2. Theory

2.1. Modality and mood

The study of modality is a frequently disputed area of research in the English grammar (Ziegeler 2020: 418). Although a vast amount of literature covering the topics of mood and modality has emerged in the recent years (Aarts and Meyer 1995: 12), there is no complete agreement in how the terms are used or as how they should be (Matthews 2014: "modality"). Definitions are nonetheless found in the literature.

Modality is a semantic concept concerned with notions such as *possibility, probability, necessity, obligation, permission, intention, hypotheticality,* and *ability*. What these notions have in common is that they are used to talk about situations that are non-factual, for example when a certain situation is not being known at the time of the utterance, but could still take place (*possibility*), or when someone is obliged to do something and they become compelled to bring it about (obligation) (Collins 2009: 11; Aarts 2011: 275). Modality and mood are closely related terms, the former is used to refer to the wider, semantic concept, and the latter to one of the grammatical sub-categories of modality (Palmer 2003: 2).

Mood is often defined as a grammatical, morphosyntactic category of the verb, like tense and aspect (Palmer 1986: 21). Traditional grammar recognizes three distinctive moods: the *indicative*, *imperative*, and *subjunctive* (Zeigler 2019: 420; Matthews 2014: "mood"), which were originally implemented in the English language via verb-inflections. In modern English, however, the inflectional system plays only a minor role in the realization of moods (Collins 2009: 11), and since the subjunctive verb form is practically nonexistent (Aarts 2011: 25–26; Palmer 2003: 3), the subjunctive is no longer regarded as a mood category in modern frameworks (Matthews 2014: "mood"). Thus it is better to regard mood in English as a non-inflectional notion and to use the term analytic mood instead, conveying that modal meanings are carried by constructions, such as the combination of a modal verb followed by a lexical verb, or clause type (Aarts 2011: 276), illustrated below.

(4) I think she [modal auxiliary verb should] [lexical verb wait] at the airport (Aarts 2011: 288).

(5) His friends should demand [subjunctive clause that he get justice] (Aarts 2011: 24).

Hence the primary means of expressing modality in English is via modal auxiliary verbs (*can, could, may, must, need, ought, shall, should, will,* and *would*) (Collins 2009: 11), though as Ziegeler (2020: 421) mentions, there are other means as well. Aarts (2011: 278–313) provides a more comprehensive list of different possibilities to express modality, such as *the modal past tense, subjunctive clauses, marginal modals (semi-modals), modal idioms, lexical modals, and hedges*, illustrated below with his examples (emphasis changed from italics to bold).

The modal past tense
(6)   If I **had** a recording of them would I be able to understand it?

Subjunctive clauses
(7)   I <u>urged</u> in my previous letter [subjunctive clause that these research staff **be** treated as their present colleagues].

The core modal verbs
(8)   I **would** not live anywhere else in England.

Marginal modals (In this paper called semi-modals.)
(9)   One place that **dares** <u>to be different</u> is Sofias's Hristo Botev.

Modal idioms
(10)   We **had better** <u>keep our feet on the ground</u>.

Lexical modality (nouns, adjectives, verbs, adverbs, and particles
(11)   So I drew the inference that the **intention** was <u>that the media *should* reproduce the programme</u>.
(12)   Survivors are **likely** <u>to experience adverse physical and psychological effects</u>.
(13)   You **have** <u>to pay for these.</u>
(14)   I thought **maybe** you'd come around.
(15)   There is **so** a Santa Claus (de Haan, 2006: 39)!

Hedges
(16)   Well we're **sort of** working towards our first preformance.

Matthews (2014: "modality") describes modality as a term that is variously applied to the grammatical or lexical indications to the kind of a speech act or the degree of uncertainty with which something is said. Furthermore, he provides a useful illustration of the concept with following pairs of examples. These pairs illustrate how an utterance can differ from another in terms of different kinds of modality.

Table 1. *Illustration of modally varying clauses* (Matthews 2014: "modality")

- He has left at once. – *a declarative*
- Leave at once! – *an imperative*

- He has perhaps left. – *uncertainty*
- He has definitely left. – *certainty*

- He can't leave. – *epistemic*
- You can't leave now. – *deontic*

- You must leave now. – *obligation*
- You can leave. – *permission*

Modality may be divided into two distinct subcategories, namely *Root modality* and *Epistemic modality* (Coates 1995: 145; Collins 2009: 21), while *Root modality* is often further divided into *Deontic modality* and *Dynamic modality* (Aarts 2011: 276; Nuyts 2014: 33; Palmer 2003: 7, Collins 2009: 23). *Epistemic*, *Deontic*, and *Dynamic* may also be called the three basic semantic dimensions of modality (Nuyts 2014: 33). This thesis follows the division into these three subcategories, discussed in more detail below. Table 2 summarizes the type of modalities.

Table 2. *Three main kinds of modal meaning* (Huddleston and Pullum 2002: 52)

i	DEONTIC	*You **must** come in immediately.*	*You **can** have one more turn.*
ii	DYNAMIC	*Liz **can** drive better than you.*	*I asked Ed to go **but he won't***
iii	EPISTEMIC	*It **must** have been a mistake.*	*You **may** be right.*

2.1.1 Deontic modality

Deontic modality is concerned with actions, both by the speaker and others, and can be defined in terms of *permission* and *obligation* (Palmer 1986: 96; Ziegeler 2020: 425). Westney (1995: 42), adds notions such as *ability*, *desire*, and *intention*, and he says (quoting Bybee and Pagliuca 1985: 63) that deontic modalities are 'agent-oriented', which means that they predicate either internal or external conditions on a wilful agent (see also Collins 2009: 22). More generally, deontic modality may indicate the degree of moral desirability of the conditions expressed in the utterance typically, but not necessarily, assessed by the speaker (Nuyts 2006: 5–6). Furthermore, deontic modals are performative in the sense that they are used to perform actions, such as permitting, laying an obligation, or to make a promise or threat (Palmer 2003: 7; Palmer 1990: 69; Aarts 2011: 276). Examples of deontic use of *must*, *may*, and *can't* by Ziegeler (2020: 425).

Obligation
(17) You **must** remember this.

Permission
(18) We **may** stay here as long as we like.

Denied permission
(19) You **can't** drive a car under the age of 18.

2.1.2 Dynamic modality

Dynamic modality generally refers to the properties and dispositions of persons, etc., that are referred to in the clause, especially by the subject (Huddleston and Pullum, 2002: 178). Palmer (1990: 36) notes that in a strict sense it is not a kind of modality at all since it is not subjective. In other words, it does not include the speaker's attitude to the factuality or actualization of the situation, which makes dynamic modality more peripheral compared to the other types of modality (Huddleston and Pullum 2002: 179). Dynamic modality is concerned with the *ability* and *volition* of the subject of the clause which contains the modal verb (Palmer 1990: 36). Aarts (2011: 277) and Ziegeler (2020: 423) conclude that dynamic modality also covers neutral circumstantial meanings, i.e., when modal verbs are used refer to timeless truths or general facts, such as in (23). Examples 20–22 by Nuyts (2014: 34–35).

(20) The garage is free so you **can** park your car there.
(21) I'll **be able to** help you in a few minutes.
(22) I'll come down for dinner soon, but I **need to** finish this letter first.
(23) The arctic hare **will** turn white in winter (Bybee 1988: 373).

2.1.3 Epistemic modality

Epistemic derives etymologically from a Greek word meaning 'knowledge, understanding'. The definition of epistemic modality is relatively noncontroversial (Nuyts 2006: 6; Coates 1995: 145). It is typically concerned with the speakers assumptions or assessment that the proposition on which a certain utterance is based is true, i.e., that it applies in the world (Collins 2009: 21; Coates 1995: 145; Nuyts 2006; 6), and it often indicates the speakers confidence or lack of confidence in the truth of the expression (Coates 1995: 145). Collins (2009: 21) adds that this likelyhood falls in a cline between weak possibility ("It may be the case.") and strong

necessity ("It must be the case."). Examples on epistemic use of *must*, *may,* and *might* by Matthews (2014: "epistemic"):

(24) He **must** surely be there by now.
(25) It **may** have been lost.
(26) The train **might** be late.

Epistemic modality encompasses semantic meanings such as *possibility* and *necessity* (Palmer 1986: 58–59; Coates 1995: 145).

2.2. Modal auxiliary verbs

2.2.1 Core modals

Traditional modal auxiliary verbs in English are *can, may, will, shall, must, ought, need,* and *dare*, including the preterite forms of the first four, *could*, *might*, *would*, and *should* (Collins 2009: 12; Huddleston and Pullum 2002: 92). These are usually referred to as either *central* or *core modals* (e.g., Quirk et al. 1985: 137) and belong to the larger group of auxiliary verbs. *Need* and *dare* have a dual nature, i.e. they may be classified and used as either auxiliary or lexical verbs (Collins 2009: 12), illustrated in (27) and (28), respectively.

Auxiliary
(27) **Dare** you talk to him?

Lexical
(28) Did you **dare** to talk to him?

Auxiliary verbs can syntactically be distinguished from lexical verbs by the NICE properties, an acronym from the following constructions: [n]egation, [i]nversion, [c]ode, and [e]mphasis. As illustrated by Aarts (2011: 68–69):

[n]	The agents *will not/won't* book the tickets.
[i]	<u>Will</u> <u>the agents</u> book the tickets?
[c]	The agents will book the tickets, and so *will* the representatives.
[e]	The agents <u>will</u> book the tickets.

Auxiliary verbs can be followed by not, or -n't (negation), moreover they can invert places with their subjects (inversion). Auxiliary verbs can be stranded, i.e., appear without a lexical verb accompanying it (code) and be used for prosodic emphasis such as in the last example above. These constructions are not possible with lexical verbs.

Additionally, Aarts (2011: 281) mentions three features that are pertained only by modal auxiliary verbs. Firstly, they are always followed by a bare infinitive verb such as in (29) or by another auxiliary verb as in (28). Secondly, modal auxiliaries are always tensed, there are no participles or modal infinitives (29). Thirdly, core modals are defective, which means they have past tense and negated forms (with the expection of *must*), but no third person singular present tense -*s* (30).

(29) There is a crisis, and he **must** act now.
(30) The agents [modal auxiliary verb will] [progressive auxiliary verb be] [lexical verb booking] the tickets.
(31) * **has mayed**, * **are musting**, * **to shall**[1]
(32) * He **cans** do it.

---

[1] Asterisk (*) marks an ungrammatical or abnormal construction.

Besides these three characteristics Huddleston and Pullum (2002: 107) mention that the first verb of the main clause in a conditional sentence (apodosis) must be a modal auxiliary, as *could* in (33). Lastly, the preterite modal auxiliaries *could*, *might*, *would*, *should* can be used in a modal remoteness meaning without the same restrictions that other verbs have. For example, in (34) *could* is ambiguous in the sense that it may indicate both past time and modal meaning, while *were* in (35) must indicate past time and not modal meaning.

(33) If you came tomorrow, [you **could** help with the flowers].
(34) **Could** you move it?
(35) **Were** you able to move it?

2.2.2 Semi-modals

As noted above, core modals form a moderately homogenous group. However, as Biber et al. (1999: 483) express, "the boundary between modals and lexical verbs taking infinitive complementation is in some cases unclear." They present two distinct groups that fall into this boundary. Marginal auxiliary verbs contain expressions such as *need (to)*, *ought to*, *dare (to)*, and *used to*, which may behave like modal verbs in some aspects, such as taking auxiliary negation and yes-no question inversion, e.g., *needn't*, *dare she do.* Such constructions are nevertheless extremely rare and largely confined to British English (Biber et al. 1999: 484).

Additionally, they list a number of idiomatic phrases that have similar functions to those of modals: *(had) better*, *have to*, *(have) got to*, *be supposed to*, *be going to*. Collins (2009: 301) describes these *modal idioms* as "…idiosyncratic verbal formations which consist of more than one word and which have modal meanings that are not predictable from the constituent parts." These expressions are called **semi-modals** (Biber et al. 1999: 484; Hansen 2018: 9)**,** a widely

accepted term in the literature[2]. This term will from now on in this paper be applied to denote both two groups described above.

Semi-modals are not a clearly defined group (Collins 2009: 15; Leech et al. 2009: 91). Ideally, according to Westney (1995: 11), a semi-modal displays three features: (A) *grammaticalization*, which requires there exist a set of both syntactic and semantic features that together indicate a grouping of some significance but with less than categorial status. Typically grammaticalization involves syntactic simplification, phonological weakening, semantic bleaching and generalization (Collins 2009: 18). (B) *idiomaticity*, that the meaning of a complex form is not simply a function of the meanings of its components. Finally, (C) *semantic relatedness*, which means that there should exist a semantic relatedness between a semi-modal and a respective core modal, e.g., *need to* and *must*.

2.2.3 Semi-modal *need to*

Quirk et al. (1985) considered *need* a marginal modal auxiliary verb and did not separate *need to* as an independent modal. Similarly, Krug (2000: 202–203) discusses both *need* and *need to* as one modal. However, *need to* has gained the status of an independent semi-modal from the core modal *need* (Leech 2003: 230; Smith 2003: 245) and hence Leech et al. (2009: 93) describe that *need* can be split into two different verbs with similar, though not identical meanings. Firstly, when constructed as a core modal auxiliary verb, *need* occurs as an invariant verb form with a bare infinitive complement. Secondly, *need*, when constructed as a lexical verb, occurs as a main verb with regular inflections before a *to*-infinitive. Nokkonen (2015: 30–32) and Hansen (2018: 9) follow this distinction and acknowledge the existence of two separate modals.

---

[2] Alternative terms denoting the same concept in the literature include *quasi-modals* used by e.g. Collins (2009), *periphrastics* used by, e.g., Westney (1995), and *emerging modals,* used by, e.g., Leech (2013) and Krug (2000).

Need has all the properties of a central modal, it has no final *-s* in the third person singular present indicative, it occurs with a bare infinitive, and it also has the contracted form *needn't*. *Need to*, in turn*,* acts as any regular transitive verb, it requires a dummy *do* for negatives and questions, a final *-s* for the third person singular, and a *to* -infinitive (Nokkonen 2015: 30–32). Moreover, need appears only in non-assertive contexts (Jacobsson 1974: 62), which include contexts such as questions, negation, shifted negation, semi-negatives, hidden negation, comparative clauses, use after superlatives, and subjunctive forms (Jacobsson 1974: 60–62). Table 3 illustrates the differences between the two modals.

Table 3. *Uses of the core modal and the semi-modal construction* (Nokkonen 2015: 31)

	NEED	**NEED TO**
Positive	-	He **needed to** come.
Negative	He **needn't** come.	Ho doesn't **need to** come.
Non-assertive contexts	He hardly **need** come.	He hardly **needs to** come.
Interrogative	**Need** he come?	Does he **need to** come?
Neg. -interrogative	**Needn't** he come?	Doesn't he **need to** come?
Code positive	-	He **needs to** come and so do I.
Emphatic affirmative	-	He does **need to** come.

The aim of this thesis is to study specifically the semi-modal *need to* and the core modal *need* is thus excluded from the analysis. When collecting the relevant data from the corpus, *need* can be excluded by adding the infinitive marker *to* in the search query, since, as discussed above, *need* only appears with a bare infinitive.

2.3. Coates' framework

According to Coates (1983: 9), indeterminacy is an innate feature of natural languages. She argues that both root and epistemic modality are fuzzy sets, something she defines, citing Zadeh

(1972: 4), as "a class in which the transition from membership to non-membership is gradual rather than abrubt." Coates thus disagrees that root modality should be further divided in the distinct categories of deontic and dynamic modality, and suggests based on her data that root meaning should be assigned on a cline from weak to strong instances, i.e., from cases that can be paraphrased with "it is obligatory" to cases that can be rephrased as "it is important," and the basic, intermediate meaning being "it is necessary for" (Coates 1983: 32).

She continues that the interpretation of strong and weak instances is further complicated by the presence or absence of speaker involvement (subjectivity), which means that the meaning of *must* may in some strong root examples become paraphrasable with "I order you to," as in an imperative clause. Nevertheless, Coates points out that such cases of subjective use of root must could be paraphrased with "it is necessary… and I order you to", moreover, there is no clear division between cases that involve subjectivity and those that do not (Coates 1983: 32–33). Hence she proposes a general model, shown in figure 1, containing the categories *core*, *skirt*, and *periphery*, which together comprise the fuzzy set (Coates 1983: 12).
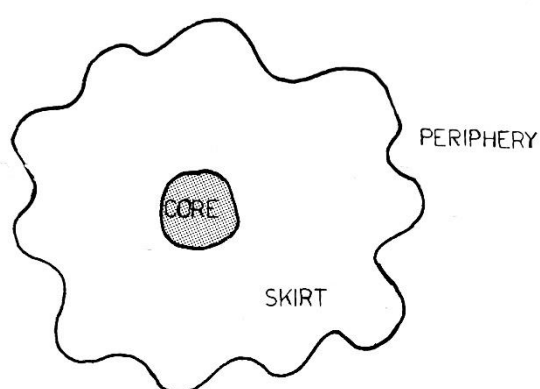

Figure 1. *Fuzzy set diagram of root **must*** (Coates, 1983: 12)

Coates (1983: 13) clarifies the grading system as follows. Modal auxiliaries that place themselves in the core group are identified as subjective in both the root and epistemic meaning, and strong in the root sense. Furthermore, the core represents the meaning first learned by children and usually corresponds to the cultural stereotype, although they are statistically infrequent. Examples identified as objective or weak settle on the periphery, and it is often possible to define such instances as being opposite to the core (not being characterised by the same properties). Examples that are intermediate in grading belong to the skirt. The majority of all examples in actual language data fall in the skirt and periphery.

In her analysis of the root meaning of *must*, a modal auxiliary of obligation and necessity that is comparable with *need to*, Coates provides a set of four features that are stereotypical and illustrate a core instance of a root use of *must* (Coates 1983: 33):

    i.   Subject is in animate.
    ii.   Main verb is an activity verb.
    iii.   Speaker is interested in getting subject to perform the action.
    iv.   Speaker has authority over subject.

She also gives an example of an instance of *must* which satisfies all the above criteria from her corpus data (Coates 1983: 34):

(36) "You **must** play this ten times over", Miss Jarrova would say, pointing with relentless fingers to a jumble of crotchets and quavers.

Furthermore, Coates has constructed an indicative matrix of eight properties which she applies when she describes the relative force of actual language use of *must* in her corpus data. The function of the matrix is more indicative than definitive (Coates 1983: 36). The matrix is

illustrated in table 4, as it would look like when analyzing the use of *must* in example sentence

(33). Here the symbols $\boxed{+}$ and $\boxed{-}$ indicate whether or not the instance that is being studied

fulfills the corresponding parameter, respectively. $\boxed{?}$ is used if the analysis of a parameter is

contrived or otherwise impossible. A strong instance of *need to* should satisfy at least five of

the first six parameters (a to f), instances in the skirt satisfy two to three of the first six, and the

periphery is represented by instances that fulfill the last two parameters (g) and (h) (Nokkonen,

2015: 85).

(a)   second person subject
(b)   subjective[3]
(c)   speaker has authority over subject
(d)   verb is agentive
(e)   paraphrasable by 'it is obligatory/absolutely essential that'
(f)   animate subject
(g)   paraphrasable by 'it is important that'
(h)   inanimate subject

Table 4. *Matrix to indicate the relative strength of deontic/dynamic (root) use of **must*** (Coates, 1983: 37)

a	b	c	d	e	f	g	h
+	+	+	+	+	+	-	-

Since *need to* is a comparable modal auxiliary with *must*, this matrix will be utilized in the

empirical section to describe the strength of the instances of *need to* in the actual data in a similar

manner as Nokkonen (2015: 88–114). This matrix is fairly transparent and most properties are

---

[3] Coates originally uses both terms *speaker involvement* and *subjectivity*. The latter was opted for use in this paper as there was more literature available to define the term.

self-explanatory. In order to be able to apply the matrix to the data as accurately as possible, the properties (b) and (d) require a closer discussion. *Subjectivity* and *agentivity* are hence briefly discussed in the following chapters. Coates unfortunately does not state how she defined and applied the terms in her study.

2.4. Subjectivity

A feature related to modal expressions is the notion of subjectivity, in other words, speaker involvement. According to Nuyts (2006: 45), this dimension is often associated with epistemic modality, since, as discussed earlier, epistemic modality is inherently speaker-oriented. In terms of epistemicity, subjective modality involves a purely subjective guess regarding the truth of the state of affairs, whereas objective epistemic modality contains an objectively measurable chance that the situation in question is true (Nuyts 2006: 45). Consider, for example, the epistemic use of core modal may in the example (37). Here, as Nuyts (2006: 45) explains, the speaker may either intend to present a merely hypothetical fact, i.e. their subjective interpretation, or a mathematically computable chance that he is married if the speaker would, e.g., have knowlegde about the community and social circles that alfred lives in.

Subjective / Objective
(37) Alfred **may** be unmarried.

Westney (1995: 44), on the contrary, says that subjectivity and objectivity are discussed in reference primarily to deontic modality. Prototypical deontic modality is also subjective, as it is usually the speaker who is the deontic source, as, e.g., when the speaker grants a permission or imposes an obligation. However, when expressing rules and regulations, deontic modality is commonly objective (Huddleston and Pullum 2002: 183), as it is something that originates from

rules or regulations. Examples (38) and (39) of subjective and objective use of deontic may, respectively (Huddleston and Pullum 2002: 183).

Subjective
(38) You **may** have one more turn.

Objective
(39) We **may** borrow up to six books at a time.

Subjectivity is sometimes claimed to differentiate core modal auxiliaries from semi-modal auxiliaries (Collins 2009: 28; Hansen 2018: 11). This idea is commonly illustrated by comparing the core modal *must* to the semi-modal *have to,* e.g. Westney (1995: 45):

(40) My girl **must** be home by ten.
(41) My girl **has to** be home by ten.

According to this view, the deontic use of *must* in the example above is more speaker oriented, as it is his or her will that the girl is home by ten. In the latter the deontic source is external to the speaker which in turn leads into an objective orientation (Hansen 2018: 12). However, Depraetere and Verhulst (2008: 23) empirically studied this claim and found out that both *must* and *have to* occur with internal and external deontic sources and that the choice of each modal did not correlate with the source of obligation.

2.5. Agentivity

According to the Oxford Dictionary of English Grammar (Aarts et al. 2014: "agentive"), verbs can be described as either agentive or non-agentive, as in examples (42) and (43). They define

an agentive verb as a verb that posits an animate instigator of the action. In other words the verb has an agent as one of its arguments.

Agentive verb

(42) The postman banged on the door.

Non-agentive verb

(43) The door was banging in the wind.

Cruse (1973: 11), nonetheless, writes that agentivity is an indeterminate concept and attempts a more critical examination. He states that the concept of agentivity is rather a relational feature than something that could be specifically applied to nouns, verbs, or clauses. Besides defining agentive verbs by the referential properties of their subject nouns it is possible to adopt a contextual approach to meaning. Gruber (1967: 943) lists three criteria that distinguish agentive verbs from non-agentive, comparing the verb *look* to the verb *see*. Firstly, all agentive verbs are substitutable by the phrase 'do something', secondly, they may be modified by a purpose phrase beginning with 'in order to', and thirdly, they might be accompanied with a manner adverbial such as *carefully*. He illustrates this as follows:

(44) John looked through the glass carefully.
(45) * John saw through the glass carefully.
(46) What John did was to look at Bill.
(47) * What John did was to see Bill.
(48) John looked into the room in order to learn who was there.
(49) * John saw into the room in order to learn who was there.

Cruse (1973: 13), however, argues that this test is unsatisfactory as it is. He prefers Halliday's (1968: 196) way of testing agentivity, in which clauses are classified according to the preferred

form of a corresponding identifying clause, these are 'do-clauses' and 'happen-clauses'. Henceworth the example (50) with verb *punched* is agentive since since it is preferrably paraphrased with a 'do-clause', and the example (53) with *broke* is non-agentive since it may be paraphrased with a 'happen-clause'. Cruse simplifies the choice between these two by saying that sentences such as (52) and (54) may be dismissed simply because they are somewhat odd, deviant, or abnormal (1973: 12).

Agentive

(50) John punched Bill.

Do-clause

(51) What John did was punch Bill

Happen-clause

(52) * What happened to John was that he punched bill.

Non-agentive

(53) The vase broke.

Do-clause

(54) * What the vase did was break.

Happen-clause

(55) What happened to the vase was that it broke.

2.6. Semantic and pragmatic functions of *need to*

Coates, as discussed above, divides the semantic field of obligation and necessity into *epistemic* and *root* modalities, the latter being a gradient phenomenon. This study will, nonetheless, utilize the division of *root* modality into *deontic* and *dynamic* modalities when discussing the semantics/pragmatics of *need to*, in correspondence with, e.g., Collins (2009: 73–77). The three categories were defined in sections 2.1.1, 2.1.2, and 2.1.3.

Nokkonen (2015: 150–153) discusses the semantics and pragmatics of *need to* in more detail. She comments that the core sense of the modal indicates a kind of inherent necessity that relates to the subject referent, similarly to the meaning of the lexical transitive verb *need* (also Leech, 2009: 109–110). This distinguishes *need to* from the core modal *must* and the semi-modal *have (got) to*, which both indicate necessity or obligation that arises from a deontic source or empirical ciscumstances (outside the subject referent), as in the example (56) (Perkins 1983: 60–62). Perkins continues that a compulsion that comes from within is always objective even with a first-person subject in the sense that the speaker has no conscious control over it. If I need to drink, I cannot control my thirst, or if I need to make a new start in life, it is because something inside me rebels the way of life I am leading (Perkins 1983: 62).

> (56) Boris **needs to** sleep ten hours every night for him to function properly (van der Auwera and Plungian 1998: 80).

Despite this core semantic meaning, *need to* has increasingly been used in contexts where the speaker recommends a certain behaviour or directly obligates the addressee (Nokkonen 2015: 151; Smith 2003: 260). In such deontic contexts there may still exist semantical differences between the different modal expressions of obligation, and these have been described in terms of subjectivity (Nokkonen 2015: 151). Consider for example the following pairs (Leech 2004: 147):

> (57) You **must** get a hair-cut.
> (58) You **need to** get a hair-cut.
> (59) She **ought to** feel wanted.
> (60) She **needs to** feel wanted.

In the first sentence, when obligating with *must*, the speaker can be assumed to use his or her authority over the addressee, whereas in the second example the situation is more complex. Here the speaker is merely pointing out that the addressees hair may look untidy, that it is too long, and that it is in the best interest of the adressee's own sake to get their hair cut (Leech, 2004: 147). A similar comparison can be made with examples (59) and (60), where *ought to* expresses an external constraint, and *need to* an internal constrait (Leech 2004: 147). In order for *need to* in (58) to be interpreted similarly to *must* in (57), there would need to exist a clear authority structure between the recipients. According to Smith (2003: 260), this ambiguity also makes it possible for the speaker to downplay his or her authority by using *need to* in place of *must*.

Furthermore, Nokkonen (2015) develops a frame for analyzing the semantic and pragmatic variation of *need to* in corpus data. She relies partially on Coates' fuzzy set theory, discussed in the previous section, but also on the works of van der Auwera (1999) and van der Auwera and Plungian (1998), who divide non-epistemic (root) modality further into the subdomains of *participant-internal* and *participant-external* modality. Participant-internal, in their terms, refers to possibility or necessity that is internal to the participant engaged in the state of affairs. Participant-external, in turn, refers to circumstances that are external to the participant and that make the state of affairs either possible or necessary (van der Auwera and Plungian 1998: 80). Morover, they explain how modal meaning evolves from lexical meaning to participant internal necessity, to participant external necessity, to deontic meaning and finally to epistemic meaning (Nokkonen 2015: 152). Figure 2 illustrates the relationship of the semantic framework and its categories developed by Nokkonen, the arrows indicate historical change in accordance to van der Auwera and Plungian (1998).
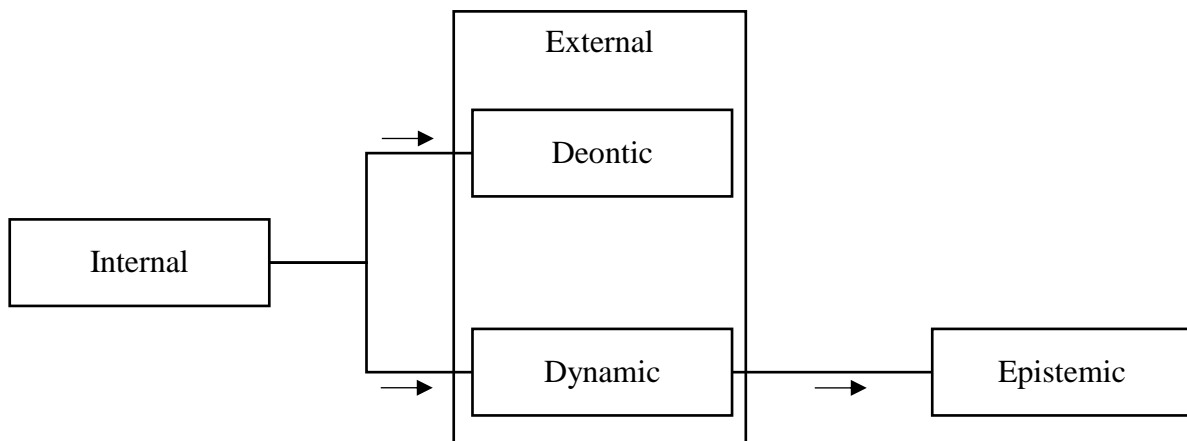
Figure 2. *The different meanings of need to* (Nokkonen 2015: 153)

Nokkonen divides the semantic/pragmatic functions into five groups, described below. All the examples are from the spoken English corpora she uses as source material, the Demographically Sampled part of the British National Corpus and the Context-Governed part of the British National Corpus).

1. Participant internal
2. Participant external (Deontic 1)
3. Participant external (Deontic 2)
4. Participant external (Dynamic)
5. Epistemic

Participant internal refers to instances in which *need to* denotes its basic lexical meaning, an internal compulsion which originates in the subject referent. Instances in this group are objective without detectable speaker control, and have only animate subjects (Nokkonen 2015: 153). In her data, the most frequent agent types are third person subjects and subjects with first person I.

Participant internal

(61) It's very important to me to get my life sorted out. I **need to** get myself back on my feet and forget all this and get it all sorted out and so I can live my life again.

The participant external field includes both deontic and dynamic instances. These express different degrees of imposed or reported obligation or necessity. They may contain a mixture of internal and external compulsion as the speaker is simultaneously appealing to the assumed needs of the addressee and obliging (Nokkonen 2015: 154). She makes a division between stronger deontic use (*deontic 1*) and weaker deontic use (*deontic 2*). In the former group the main verbs are typically agentive activity verbs, and the subject is most commonly *you*. The instance is stronger if there exists an authority structure between the speaker and the addressee.

Deontic 1

(62) Oh well that's what it is then, yeah. Yeah okay. <pause> But you'll **need to** sort that out with Mary when she's back, she's not back till after Easter anyway.

In the latter group there are weaker deontic instances that often have mental and existential main verbs, although agentive verbs are still frequent. The subject is often a generic *we*, but a generic *you* is also frequent as are passive constructions. The directives in this group tend to be more hidden and hedged (mitigated), and they are aimed at at a more vague group of recipients (Nokkonen 2015: 154).

Deontic 2

(63) In this society of ours, many children are led to confusion by the complexity of the life in which they're placed. And I think for the dyslexic child, for the disturbed child generally, we **need to** offer an atmosphere which is calm.

Dynamic uses of *need to* express only vague external necessity arising from circumstancial factors or the required qualitites of the subject referent in a situation, not from regulations or obligations imposed by an authority (as in a *deontic* case) or from an internal necessity experienced by the subject, as in the *participant internal* cases. Dynamic instances typically have an inanimate subject or existential *there* as a subject. The main verbs are existential or occurrence verbs that do not express any volitional activity.

Dynamic

(64) There **needs to** be er a banking commission to take these functions away from a Bank of England <unclear> performed so poorly er and was shown to have done by the er b--by the bingham report.

Epistemic group includes the cases where *need to* is used in epistemic sense. Nokkonen uses the term epistemic in its standard meaning. According to Coates (1983: 44), epistemic use of the core modal *must* commonly occurs with syntactic features such as existential, inanimate subject, stative verb, perfective and progressive aspect, adding that these features are common to all modals expressing epistemic sense. Nokkonen comments that this also applies for the few epistemic instances of *need to* in her data. Since these syntactic features are similar to those of a dynamic meaning, the borderline between a dynamic and epistemic instance becomes unclear (Nokkonen 2015: 155). She thus argues that *need to* is sliding through the dynamic domain into epistemic meaning (as illustrated in figure 2).

Epistemic

(65) Now obviously in a case like that there doesn't really **need to** be a
great deal of negotiation because obviously if its something being
given ex gratia the really obviously people really cant be in a
position to argue too much.

2.7. Previous research

Leech (2003) has studied the development of both core modals and semi-modals in American
and British English between the 1960s and the 1990s. He used six different corpora in his study,
the British *LOB* and *F-LOB* from 1961 and 1991, and their respective counterparts, the
American *Brown* and *Frown* from 1961 and 1992, in addition with two corpora of spoken
British English, the *SEU-mini-sp* including the years 1959–1965, and the *ICE-GB-mini-sp* from
1990–1992. The core modals Leech included in his study were *would, will, can, could, may,*
should, must, might, shall, ought (to), and *need(n't).* Additionally, he discussed some semi-
modal expressions, namely *be going to*, *gonna*, *be to*, *(had) better*, *(have) got to, gotta, have*
to, need to, want to, *wanna*, and *used to*.

Leech's study explored the change of frequencies of each modal in the chosen corpora. Leech
observed that the infrequent core modals *shall*, *ought to*, *need,* and the middle-frequency
modals *may* and *must* had decreased drastically. *Would* and *should* had declined less
dramatically, and the modals *will*, *can*, *could*, *might* had not changed significantly. The overall
use of semi-modals is increasing, particularly *need to*, which had increased remarkably in both
American and British English. All the semi-modals are nonetheless much more infrequent than
the core modals. Results in the spoken corpus data confirmed the decrease of core modals and
increase of semi-modals. Finally, he explored the semantic aspect of three declining modals,
may, *should*, and *must*. The results showed that the dominant senses of *may* and *should* in the
1960s became even more dominant in the 1990s, while the minor senses became even more

marginal. Leech argues that this suggest a tendency for declining modals to become more monosemous, which means that one sense dominates over others. This did not apply to *must*.

Smith (2003) has explored recent developments in the behavior of core modals *must* and *need*, in addition with semi-modals *have to*, *(have) got to*, and *need to* in predominantly written British English. His aim in the study was to discover whether there had been significant shifts in the frequency of the respective forms and their associated meanings. The modals he included share a common semantic element, they are all modals of strong obligation or necessity, and to some extent also epistemic necessity. Smith used the same corpora in the analysis as Leech (2003: 223–240), *LOB, F-LOB, Brown, Frown, SEU-mini,* and *ICE-GB-mini*. His results revealed that the different modal markers of obligation and necessity fluctuated significantly between 1961 and 1991.

Must is only thriving in its epistemic sense. Its use in its root sense is declining, most likely because it is prototypically subjective, insistent, and sounds authoritarian, which makes it unfavorable in a society where overt marks of hierarchy are avoided. Smith notes that *have to* has at best only partially filled this void left by the decline of must, and hence *need to* has stepped in. The frequency of *need to* had increased more than threefold in three decades, and interestingly at a similar rate in both British and American English. Most likely on account of its different core meaning to the other modals of obligation, and its potential to be used as an indirect means of imposing obligations. Smith concludes that there was no clear evidence for American influence on British English other than that of the declining core modals *must* and *need*.

Collins (2009) has conducted a more comprehensive study on the English modal system. He explored modal expressions found in three parallel corpora, the British and Australian components of the *International Corpus of English (ICE)*, and a specifically compiled American English corpus consisting of the spoken component of the *Santa Barbara Corpus* (recorded in the 1990s) and, for the written part, texts chosen from the *Frown corpus*. Collins analyzed every token (46,121) of the core and semi-modals found in the data and divides the analysis of each modal auxiliary into different modal meanings, also considering temporality, negation, and regional and stylistic variation. He concludes that *need to* is semantically alike with the modal *need*, and that American English is leading the way in the recent rise in its usage. He proposes that this is due to the attractive option *need to* provides in its deontic use, it enables the speaker to formulate a requirement that at the same time acknowledges and endorses the subject-referent's needs, something that is not expressed by any other expression of deontic necessity (Collins, 2009: 161).

Johansson (2013) has studied the core modal *must* and semi-modals *have to, have got to*, and *need to* in the *Corpus of Contemporary American English* (COCA). The focus of his research was how the frequencies of these modals changes in COCA, which spans from 1990 to 2005, and how their distribution varies between different genres. He compares the results with data retrieved from another corpus, *the Time Magazine Corpus of American English*. He notes that although the picture is crude, the findings are in line with previous studies, in particularly those of Leech (2003) and Smith (2003). The frequency of *need to* has risen steadily over the past years. *Need to* is also more common in the spoken genre in comparison to *must*, although in the fiction genre the relationship is the other way around. Moreover Johansson came across a pattern in a related use of the lexical verb *need*: I *need* you to + verb. For example, *I need you to* sign this paper. For Johansson this represents a new strategic way of asking someone to do

something, an appeal instead of a direct command. He concludes the paper with a cross-linguistic perspective, contrasting the chosen modals to corresponding expressions in Norwegian, commenting that English seems to have a more varied repertoire for the expression of obligation.

Nokkonen (2015) has studied the contemporary use of the semi-modal *need to* in British English by using corpus-based methods and sociolinguistics. The corpora were chosen to enable the exploration of the semantic and pragmatic variation of need to across numerous sociolinguistic variables, including real time, medium, age, gender, social class, and several spoken registers. She used following corpora, the British *LOB* and *F-LOB*, *The London-Lund Corpus of Spoken English* (LLC), *The Bergen Corpus of London Teenage language* (COLD), *The Demographically Sampled part of the British National Corpus* (DS), and *The Context-Governed parts of the British National Corpus* (CG). For the purposes of her analysis she created a framework that categorizes the different functions of need to in accordance to its semantics. She calls these categories *participant internal, participant external (deontic 1, deontic 2, dynamic),* and *epistemic.*

Her main findings concerning the semantic and pragmatic variation of *need to* was that it depends most of all on the subject, both person and type. Another important factor is the authority structure in the speech situation. Most instances of *need to* in her corpus data fall into the non-epistemic categories, although there were also a few epistemic instances. The study of the real time variable disclosed that the frequency of *need to* had risen significantly in both spoken and written language in a time span of thirty years, and that it is used less in written than in spoken language. Its use gradually decreases and finally drops from young adults to the oldest age group, who also retained the more traditional uses. The results were indicative that

in interactive private and spoken registers *need to* was favored by men, and that men employ the more subjective functions that require taking stance, while women generally applied more the internal function. *Need to* was overall mostly favored by upper class people. Nokkonen concludes that the social and linguistic patterns governing the use of *need to* can be highly complex and that the traditional concepts and frames can't be applied in a rigid way.

Glass (2015) has written an article on *need to,* investigating the social reasons why a speaker might favor *need to* on the expense of *have to* or *got to*. Her focus was especially on the second person use *you need to*. She argues that by choosing *need to*, the speaker unambiguously signals that he or she is aware of the hearer's needs and licensed to tell him or her what is best for them. Thus, Glass hypothesizes that *need to* is more commonly used by speakers with a mentoring role or authority over the hearer, or those with more knowledge about the relevant domain. She investigated a set of corpora, the Providence section of *The Child Language Data Exchange System* (CHILDES), *The Michigan Corpus of Academic Spoken English* (MiCASE) which contain relevant metadata about the how the speaker and hearer relate to each other, concluding that the corpus-data confirms the hypothesis.

In a more recent study on the patterns of modal development, Daugs (2017) researched the development of modals and semi-modals in American English using two large corpora, COCA and *The Corpus of Historical English* (COHA). Since most of the earlier studies on the subject are based largely on the Brown family of corpora, Daugs measures his data against earlier findings, namely the overall decrease in the use of modal verbs, the persistence of the trend, the underlying pattern observed, and the rise of the semi-modals. His results concluded that while the overall decrease of modal frequency seems irrefutable from the 1950s onwards, the evidence is less conclusive for the 19th and early 20th century, as the frequencies fluctuate over

the course of this period. The high-frequency modals (*would*, *will*, *can*, *could*) do not behave homogenously, the demise of *will* being responsible for over one-fifth of the overall losses of the modals. Comparing the trajectories of the modals, *need to* seems to have reached the same phase of grammaticalization which the modals *have to*, *want to*, and *be going to* went through about a 100 years earlier. Despite these developments the semi-modals are still seriously outnumbered by the core modals in all registers and Daugs finds this lack of competition intriguing.

3. Methods and materials

3.1. Corpus linguistics

Crystal (2008: "corpus") defines *corpus* as a collection of linguistic data that consists of either written texts or transcriptions of recorded speech, something that can be used as a starting point of linguistic description or to verify hypotheses about a language. A corpus that is designed for linguistic analysis is normally a systematic, planned and structured compilation of texts, which separates it from a text archive, which in turn is a text repository that is normally opportunistically collected and not structured (Kennedy 1998: 4).

The term *corpus linguistics* was first used in the 1980s, although it is generally agreed that this sub-discipline in language studies existed already in the 1960s (Leech et al. 2009: 24). Biber and Reppen (2015: 2), however, remark that the standard practice in linguistics until the 1950s was to base language descriptions on analyses of collections of natural texts, in other words, pre-electronic corpora. They further point out that some dictionaries and grammars were also based on the analyses of natural texts before the 1950s. Major linguistic studies based on electronic corpora began to appear in the 1980s as electronic corpora and the computational tools that made possible to analyze those corpora became more easily available (Biber and

Reppen 2015: 3). Corpus linguistics has become the mainstream paradigm in the study of languages particularly in the English language, where increasingly rich and varied corpus resources have become available (Leech et al. 2009: 8).

Biber, Conrad, and Reppen (1998: 4) describe the essential characteristics of corpus-based analysis. Firstly, it is empirical, i.e., it analyses the actual patterns of language use in natural texts. Secondly, it utilizes a large and principled collection of natural texts as the basis of analysis. Thirdly, it makes extensive use of computers for analysis, using both automatic and interactive techniques, and finally, it depends on both quantitative and qualitative analytical techniques. Together these characteristics result in a scope and reliability of analysis that would not be achievable otherwise (Biber et al. 1998: 4). Although it is impossible to study languages in their entireness, a well-defined and compiled corpus may provide a moderate representation of a domain of language use, and thus generalizations can be made. More importantly, by looking at large samples of language data, it is possible to collect enough examples of a small element of language, such as a rare modal auxiliary, and to be able to empirically study, e.g., their behavior or variation. Hence corpus linguistics is a useful tool for studying the semi-modal *need to*.

Tognini-Bonelli (2001: 17) describes two different approaches to studying corpora. The first of these is the *corpus-based* approach. This would entail that the scholar starts off with a set of explicit rules, for example that the word *any* is used with negative, interrogative, and conditional structures, and would then attempt to validate and quantify these statements with corpus data. The data itself would not invalidate these statements but would provide a quite different picture if given the chance to speak for itself (Tognini-Bonelli 2001: 17). Collins (2009: 5) formulates that in this approach the corpus is more importantly the source of

frequency data, which may influence hypotheses applied to it, or formulated based on it. The second approach is called *corpus-driven*. In this approach, theory is being built one step at a time in the presence of the evidence. Patterns are observed, which leads to a hypothesis and to the generalization in terms of rules of usage, finally leading to a unification in a theoretical statement (Tognini-Bonelli 2001: 17).

A common way to utilize corpora in linguistic research is to analyze concordances. A concordance search is a formatted version of display of all the occurrences or tokens of a particular keyword[4] in a corpus. Concordance lists may be produced in several formats. The most common format is the *Keyword in Context* (KWIC) *Concordance*, which lists all instances of the searched keyword and presents them in predetermined amount of context. In a typical KWIC concordance list the token or keyword is in the middle, its immediate context preceding and following it. This enables the semantic profile analysis of a certain token. Table 5 gives an example of how a typical KWIC list looks like.

Table 5. *A fragment of an unsorted concordance for* **on** *from the Brown Corpus* (Kennedy, 1998: 251)

Text	Context before	Keyword	Context after
2436	Let us speculate a little	*on*	the maximum size of the python
2449	and there is nothing at all	*on*	the amethystine python
2468	data	*on*	the boa constrictor about match
2476	the following information	*on*	snakes varying greatly
2482	United States, could supply data	*on*	the maturing period of
2508	amount of agreement	*on*	some of the giants.
2512	There are three levels	*on*	which to treat the subject
2514	proof and therefore may err	*on*	the conservative
2521	The third level leans	*on*	a belief that a lot of smoke
2536	but	*on*	the third level, and is chiefly
2544	Detailed information	*on*	record lengths of the giants
2548	as far as possible, data	*on*	these aspects of growth

---

[4] Sometimes referred to as types, search items, node words, or target items.

3.2. The 2006 Corpus of American English

The 2006 Corpus of American English (AmE06) is a corpus developed at the University of Lancaster by Amanda Potts and Paul Baker. It is an extension to the Brown family of corpora (table 6) and more specifically a companion corpus to *The 2006 Corpus of British English* (BrE06) corpus (Potts and Baker 2012: 301). The texts sampled in the corpora have been published between the years 2004 and 2008, their frequencies as follows: 2004 (1 text), 2005 (48 texts), 2006 (400 texts), 2007 (45 texts), and 2008 (6 texts) (Potts and Baker 2012: 302). The corpus consists of 1,175,965 words sampled from the 500 texts. As the other corpora in the Brown family, the AmE06 is divided into four subcategories: *press*, *general prose*, *learned*, and *fiction* to be representative of the American written register (table 7). All the texts that were included in the corpus had originally been published as a hard copy before being placed online. 175 of the texts were sampled from the beginning of a longer text, 125 from somewhere in the middle, and 40 from the end to prevent the data only representing the start of texts.

Table 6. *The Brown family of corpora* (Leech, 2013: 97).

	$1901^{\pm3}$	$1931^{\pm3}$	1961	1991, 1992	$2006^{\pm2}$
British English	BLOB-1901	BLOB-1931	LOB	FLOB	BrE06
American English	(no corpus yet)	B-Brown	Brown	Frown	AmE06

Table 7. *The genres of the Brown family of corpora, and the number of text samples of 2000 words each contains* (Leech, 2013: 98).

Identifying letter		Genres	Subcorpora	Text samples
	A	Press: reportage	**Press**	**44**
	B	Press: editorial		**27**
	C	Press: reviews		**17**
	D	Religion	**General**	17
	E	Skills and hobbies	**prose**	36
	F	Popular lore		48
	G	Belles letters, biography, memoirs etc.		75
	H	Miscellaneous		30

J	Learned (academic writing)	**Learned**	80	
K	General fiction	**Fiction**	29	
L	Mystery and detective fiction		24	
M	Science fiction		6	
N	Adventure and western fiction		29	
P	Romance and love story		29	
R	Humor		9	
Total			500	

Potts and Baker defined an American text as something that was written either by an author born in America and/or someone who had lived in the United States continually for most of their lives (Potts and Baker 2012: 302). The corpus is freely available online on the CQPweb (Corpus Query Processor) website (https://cqpweb.lancs.ac.uk/ame06/), which is administrated by Andrew Hardie from the University of Lancaster. It has a built-in query interface, and no external programs are needed to make searches from the corpus.

Lastly, the AmE06 corpus includes grammatical tags, which means that each word in the corpus is marked corresponding to which part of speech (POS) it represents, i.e., whether the word is a noun, verb, adjective, adverb etc. Table 8 illustrates the Oxford Simplified Tagset can be used on CQPWeb. The tagging process is automatized and done with a program called the Constituent Likelihood Automatic Word-tagging System (CLAWS), developed at Lancaster University in the 1980s. The AmE06 uses the CLAWS7 tagset, and according to the CLAWS website (http://ucrel.lancs.ac.uk/claws/) it has an accuracy rate between 96–97%. Table 9 demonstrates how tagged text data looks like.

Table 8. *The Oxford Simplified Tagset*

A, ADJ	adjective	INT, INTERJ	interjection
N, SUBST	noun	PREP	preposition
V, VERB	verb	PRON	pronoun
ADV	adverb	$, STOP	punctuation
ART	article	UNC	other / uncertain
CONJ	conjunction		

Table 9. *An extract from Bram Stoker's Dracula (1897) tagged with the CLAWS7 tagset*

Enter_VV0 freely_RR and_CC of_IO your_APPGE own_DA will_NN1 !_! "_" He_PPHS1 made_VVD no_AT motion_NN1 of_IO stepping_VVG to_TO meet_VVI me_PPIO1 ,_, but_CCB stood_VVD like_II a_AT1 statue_NN1 ,_, as_CS21 though_CS22 his_APPGE gesture_NN1 of_IO welcome_NN1 had_VHD fixed_VVN him_PPHO1 into_II stone_NN1

3.3. Units of analysis

This study focuses on the semi-modal auxiliary verb *need to*. The first step of the analysis is to retrieve all the relevant instances from the AmE06 corpus, a simple process with the built-in query interface on the CQPweb website. The search with the lemma "**need**" retrieves all the possible forms of the word *need* from the corpus, including the verb forms *need, needs, needed, needing*, as well as the noun *need*. By adding the POS tag "V" into the search query, i.e., "**{need/V}**", the noun forms may be excluded from the search. Finally, to restrict the instances to those where the verb *need* functions as a semi-modal auxiliary verb, and exclude the core modal *need*, the infinitive marker "**to**" is added to the query. Hence the final search string reads as "**{need/V} to**".

However, as mentioned before, the automatized POS tagging is not flawless. It is possible that the automatic tagging process has resulted in the verb *need* to be incorrectly marked as something else. Therefore, it was decided to run additional searches in the corpus with the

chosen search string, alternating the POS tag following the different options shown in table 8, e.g., "**{need/N} to**", "**{need/ADV} to**", "**{need/PRON} to**" and so forth, to see if there would be any instances with incorrect tagging. The search with "**{need/N} to**" returned 49 instances from the corpus. Among these results there were three instances where *need* indeed functioned as a verb but was incorrectly tagged as a noun, and thus these three were added to the analysis. Applying the other tags would return no results.

Another problem occurs with the past tense form *needed* and the corresponding participial adjective *needed*, which may also be followed by the infinitive marker *to*, as exemplified in (66) and (67). The automatic tagging is unable to distinguish these two forms from each other since both are morphologically past tense forms of the same verb. There is no workaround, so the participial forms are excluded manually from the results. This is feasible because of the small size of the AmE06 corpus and the resulting small number of tokens that require analysis.

(66)   The amount of work [participial adjective **needed**] [infinitive clause **to** finish the project].

(67)   She [lexical verb **needed**] [infinitive clause **to** finish the project].

3.4. Reliability

This study will use descriptive statistics to present the quantitative frequency data. The results are then compared to results found from other corpora in other studies, wherever possible. A simple significance test will then be applied to validate the observed differences and to rule out the possibility of a coincidence. This test is called the log-likelihood (LL) test, which can be used to study whether an observed difference in the frequencies of the same keyword in two different corpora is statistically significant.

There is access to a wizard, created by Paul Rayson, on the University of Lancaster website, that allows to perform a log-likelihood test by simply inserting the values on an online form. The form is available on http://ucrel.lancs.ac.uk/llwizard.html. Table 10 illustrates the wizard with example values added. 50 000 and 75 000 are the absolute sizes of the corpora, 52 and 57 the absolute frequencies of the keyword in the corpora, respectively. A search with these values would result in a matrix shown in table 11, in which O1 and O2 are the observed frequencies, %1 and %2 the relative frequencies of the keyword in the corpora. The $\boxed{+}$ symbol indicates overuse in O1 relative to O2, a $\boxed{-}$ would indicate the opposite.

The LL value tells whether the result can be treated as significant, the higher the value, the more statistically significant the difference. The LL score has to be above 3.84 for the difference to be significant at the 95% level, i.e., that there is a less than 5% chance that the result is a mere coincidence. In the example provided below the result is thus not statistically significant.

Table 10. *An example of the log-likelihood wizard on the University of Lancaster website*

	Corpus 1	Corpus 2
Frequency (e.g. word)	52	57
Corpus size	50 000	75 000

Table 11. *An example of a results matrix produced by the log-likelihood test wizard (a non-significant result as LL < 3.82)*

Item	O1	%1	O2	%2	LL
Word	52	0,10	57	0,08 +	2,65

4. Results

4.1. Initial numbers

The initial search with the query '**{need/V} to**' in *The American English '06 Corpus* on CQPweb returns a total of 288 instances. Table 12 includes 8 randomly picked examples from the results. Here only a little context is included, for the purposes of the analysis a larger part of the immediate context was taken into consideration. It is possible to expand the context and access the text metadata on the CQPWeb website. All the highlighting in the text examples provided from now on has been added subsequently to make it easier for the reader to follow the discussion and analysis. The genre of the text extract is provided in italics after each example, in addition with the text label letter and number, e.g., (*Fiction*, K07).

Table 12. *Example of a concordance list retrieved from the AmE06 corpus*

	Text	Context before	Keyword	Context after
1	G62	…get back to the base. We	**need to**	go now. We 're hit, we…
2	B25	…right about one thing: We	**need to**	do more than restore the prewar…
3	G03	…and they only	**needed to**	wait for a home delivery on the…
4	J21	…knowledge about what we	**need to**	do in order to learn and remember…
5	F45	… FERPA guidelines you	**need to**	consider…
6	L13	…just so I 'd be blamed. We	**need to**	make it disappear. Whoever did it…
7	B15	…certainly worsen. Georgia	**needs to**	implement every form of…
8	A19	… in 2006. Educators	**need to**	do a better job preparing boys for…

After analyzing all the 288 instances individually there were a total of 42 cases where *need* was tagged as a verb and was followed by the infinitive marker *to*, but where it functioned as a participial adjective as described in example (66). These were omitted from the analysis. One instance was unclear and left out since it appeared within a citation that referred to a poem written in English vernacular (68).

(68) We want to see you again as ruler of your own space Big Negro Big ol Negro growin wind storm flyin thru your huge blue lung Lung filled with hurricanes of transparent fingerpops and **need to** be changed to moans Stretch out negro Grow "Gro Gwan" Gro Grow Stretch out Expand. (*Learned*, J67)

In the additional searches, as noted earlier in section 3.3, there were three instances found, which were added to the final number. Thus 248 valid instances remain. The AmE06 corpus contains a total of 1,175,965 words, and so the normalized frequency (per million) of *need to* in this corpus is 211 when rounded up to the nearest integer [(248/1175965) * 1000000 ≈ 211 pmw]. Figure 3 shows the frequency along with results found from other corpora.
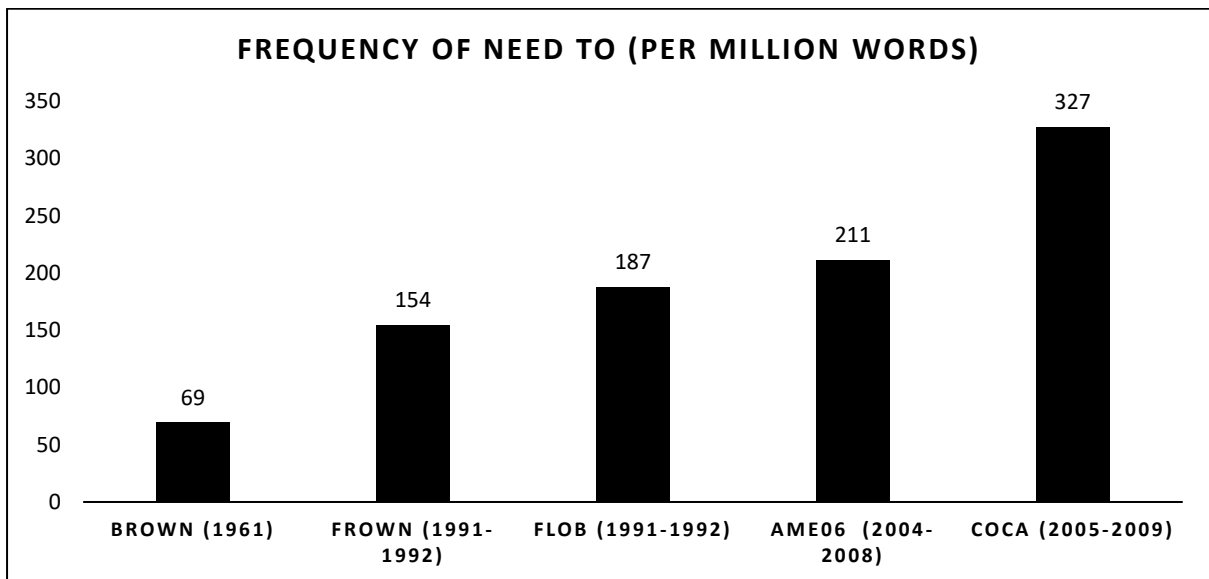
Figure 3. *Frequency of **need to** per million words in Brown (Leech 2003: 229), Frown (Leech, 2003: 229), FLOB (Nokkonen, 2015: 88), AmE06, and COCA (Daugs, 2017)*

The frequency of *need to* in AmE06 falls between FLOB and COCA. FLOB includes 187 instances per million words (Leech 2003: 229, whereas in COCA there is a total of 327 instances per million words (Daugs, 2017) in written texts. FLOB and AmE06 are compiled similarly, so the numbers are comparable. The LL test returns a value 1.55 when testing AmE06

against FLOB, hence the difference is not statistically significant. On the other hand, AmE06 compared to FROWN returns the number 9.52, which is clearly above the 3.84 threshold.

The partition of COCA considered here is published between 2005–2009 and contains roughly 80,000,000 words of American English, vastly more than AmE06. Furthermore, COCA is sampled differently from the corpora in the Brown family, so the comparison between the two needs to be treated with caution. Unfortunately, there are no raw frequencies available in the article written by Daugs (2017), so it is impossible to perform an LL test to the results. However, as the difference between the frequencies of *need to* in AmE06 and COCA is larger than that between Frown and AmE06, it should be safe to assume that this difference is also statistically significant. These results concur with the phenomenon of rising frequency of *need to* in the modal system in American English.

4.2. Distribution of subject types

Need to appears in the corpus data with different subject types, first-person singular and plural, second person, animate and inanimate 3rd person subjects, and existential *there.* The most frequent subject type in the AmE06 corpus is the third person animate subject with a total of 96 occurrences (69). Metonyms such as "the DCF" or "the school district" were counted as animate subjects (70). There was a s single occurrence of need to which had existential there as subject (71). Inanimate subjects and existential subjects were included in the non-intentional group since they cannot be obliged by the speaker. As discussed earlier, the lack of subject selection is one of the semantic criteria that defines a modal, and in the data *need to* accepts passive constructions (71), inanimate subjects (72), and existential subjects (73). This is in line with the idea of grammaticalization of *need to* which is argued to be moving towards central auxiliaries. The absolute frequencies are provided in figure 4.

(69) As the party in control of the government, <u>Republicans</u> ***need to*** convince voters that President Bush and Congress are doing a good job, that the war in Iraq was the right thing to do and that the economy is on the upswing - the issues that are driving voter attitudes. (*Press*, A06)

(70) At the same time, <u>the DCF</u> **needs to** find the right oversight balance for the foster caregivers -- as in vigilance but not oppression. (*Press*, B02)

(71) The wise founders knew that, to bring this destiny to fulfillment, <u>the work of man</u> would **need to** be completed <u>by a higher power</u>. (*General prose*, F03)

(72) <u>The transmission</u> **needs to** know the vehicle 's over-the-road speed to know when to shift gears properly. (*General prose*, E16)

(73) Moreover, <u>there</u> **need to** be massive private and public efforts to retrain workers for new jobs. (*General prose*, G25)
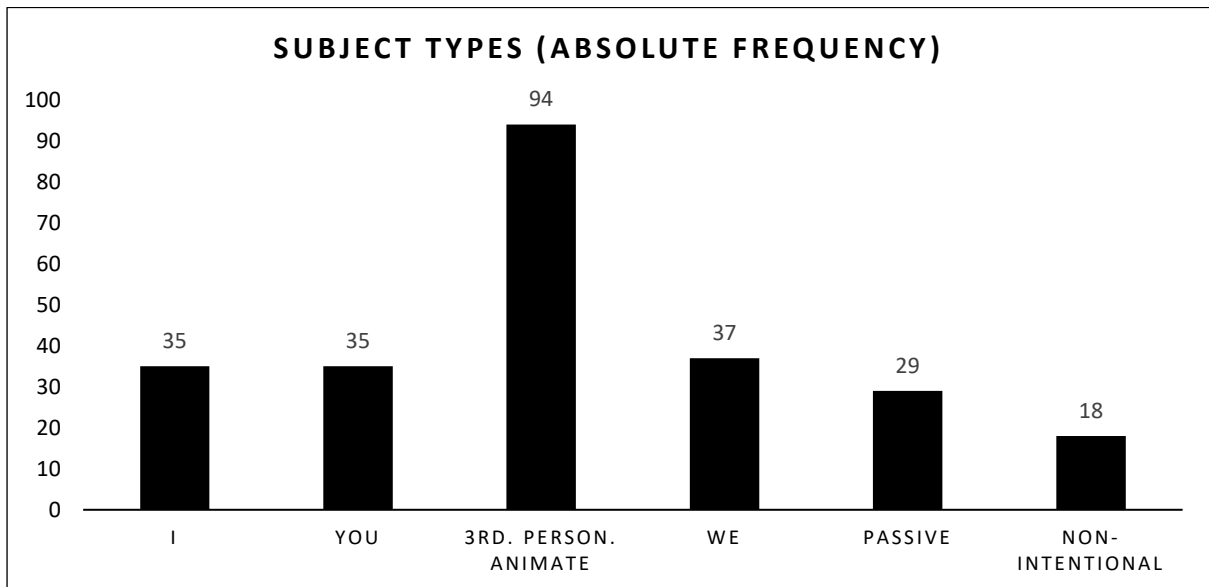

Figure 4. *Distribution of subject types of **need to** in the AmE06 (N = 248)*

The log-likelihood wizard was used to test the reliability of the figures. The LL value when comparing the frequency of the third person animate to the second most frequent subject type, *we* is 25.65, so the third person animate group is reliably the most frequent. However, the difference between *I*, *you*, *we*, and passive forms would return no significant differences, so it can be said that *need to* is evenly distributed among all these subject types. The LL value between non-intentional subjects and passive form is likewise insignificant, however when compared against first-person singular subject the value is 5.55, hence it can be deduced that non-intentional subjects are statistically less frequent than other subject types, with the exception of the passive form.

Nokkonen (2015: 88) has counted the subject types in FLOB. The comparison between the frequencies in AmE06 and FLOB is shown in figure 5. As noted earlier, the statistical difference between the total frequencies in the two corpora is insignificant. There is, nevertheless, some statistically significant variation when each subject type is looked at individually. The third person animate form is statistically more frequent in AmE06 and the passive form in FLOB. Nokkonen does not provide exact numbers how the instances with a third person animate subject vary between different genres, so it is difficult to say where the difference comes from. She adds, however, that the frequency of third person animate subject has doubled between LOB and FLOB, and that the difference is mainly in the non-fiction genres of the corpora. Regarding the passive voice Nokkonen (2015: 97–98) says that in written British English there is a major change in the academic subcorpus, there being a total of 10 instances in FLOB (compared to two in LOB). In AmE06 there are mere two instances in the academic genre, so no similar trend in American English can be deduced from the data (there is no data available from the Brown and Frown corpora). Most passive instances in the AmE06 are from the general prose section. The overall distribution in the two corpora is, nonetheless, similar.
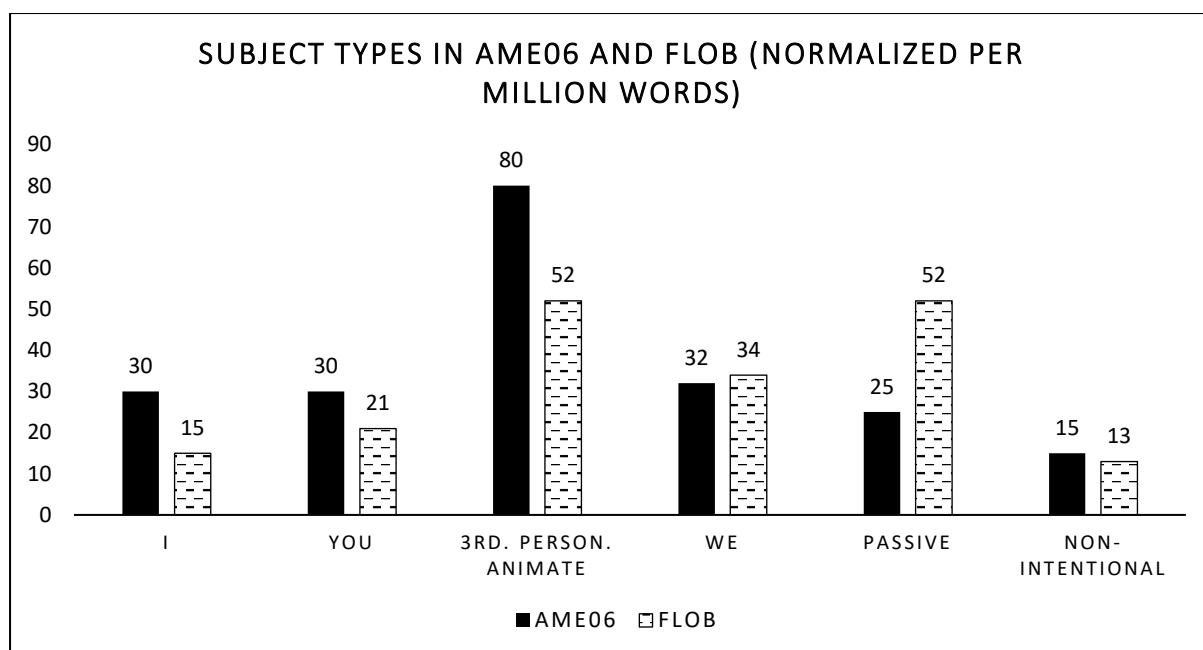
Figure 5. *Distribution of subject types of **need to** in the AmE06 and FLOB*

4.3. Subcategories

Figure 6 exemplifies the distribution of *need to* between different subcategories which each represent a different genre. This figure is only an approximation since it is based on the Brown family of corpora standard, according to which the length of every text extract is exactly 2000 words, which makes the total count 1,000,000 words. This is, however, not the case in AmE06 which contains 1,175,965 words. In this corpus the length of a single text extract varies between 2000–2400 words, and hence the exact number of words in each genre or subcategory would be required to be able to provide accurate statistics. This information was unfortunately unavailable in the corpus metadata that is provided on the CQPWeb website, or on the Lancaster University website. The percentages shown in figure 6 were counted with the approximation that each text in AmE06 consists of exactly 2000 words, hence this figure is only approximate.

Genre distribution in Frown was unavailable in literature, but Smith (2003: 252) has provided raw numbers of genre distribution of *need to* in the FLOB corpus, and these numbers were added in the figure to serve as comparison. He also notes that in FLOB *need to* seemed to have no clear genre preferences. Similarly, in AmE06 *need to* is spread along all the subcategories. As the values concerning Ame06 are indicative and not precise, no statistical tests were performed. The most formal subcorpus, *learned*, has the least instances (16%) contrasting with the most informal corpus, *fiction*, which has the most occurrences (38%). This in concordance with the idea that *need to* is still an emerging expression of modality, and change in language is driven forward primarily by the spoken register, which is more informal by nature. Fiction contains more direct dialogue which, to an extent, resembles spoken language, and in press texts it is more common that the writing contains direct quotations of people saying things. However, as there is no data available from the Frown corpus, such progression is speculative.

DISTRIBUTION OF *NEED TO* BETWEEN SUBCORPORA IN AME06

FICTION 38 %
PRESS 25 %
GENERAL PROSE 21 %
LEARNED 16 %

DISTRIBUTION OF *NEED TO* BETWEEN SUBCORPORA IN FLOB
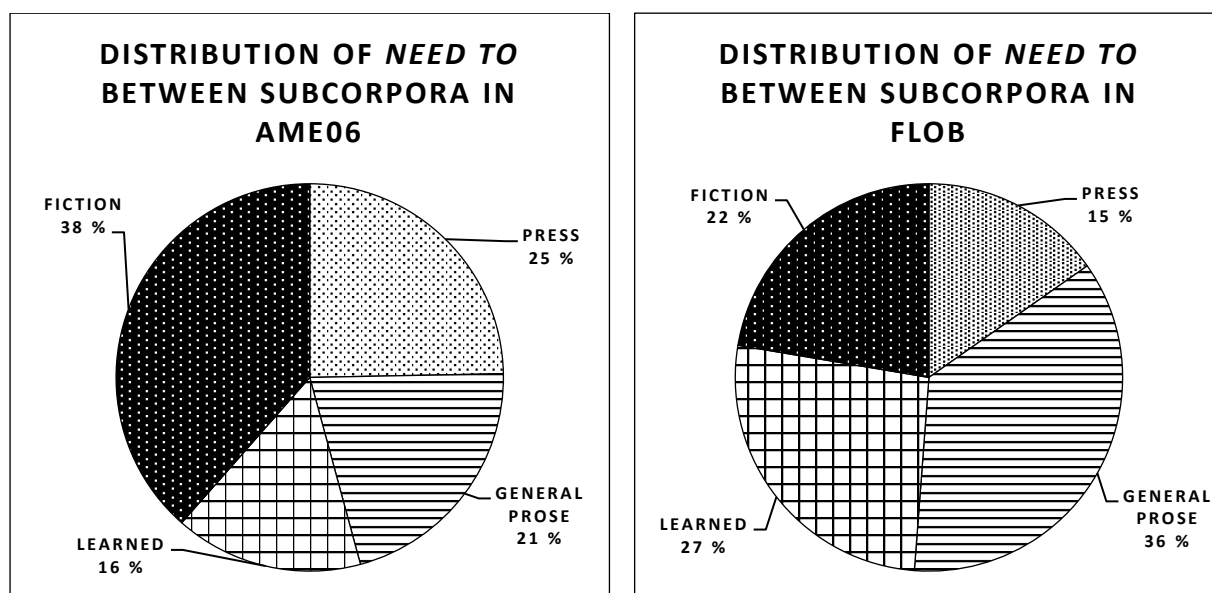
FICTION 22 %
PRESS 15 %
GENERAL PROSE 36 %
LEARNED 27 %

Figure 6. *Distribution of **need to** between different subcorpora in AmE06 and FLOB*

4.4. Context and modifiers

Need to occurs dominantly in affirmative contexts in the corpus data, a total of 90% of the instances. Negations are slightly more common (6%) than interrogatives (1%) and other contexts (2%). Figure 7 illustrates the portions. The LL test was applied to compare the number of negative to other contexts, and the resulting value was 4.72, thus the difference is statistically significant. The difference between interrogatives and other contexts was insignificant. There were no results from a similar analysis available in the literature besides Nokkonen (2015: 50), who counted the context types in corpora of spoken British English. The proportions in her data are similar to AmE06.

(74) is an example of a typical case of need to in an affirmative context. There were a few clauses in which negation was formed with an auxiliary do (75), one instance where a modal auxiliary was used to create negation, and a single instance were negation was formed with no-negation (76). Interrogatives were similarly rare, and the few occurrences were all formed with an auxiliary do (77). Other non-assertive contexts included occurrences such as (78), where the negation is hidden. one example (79) where *need to* appears in a comparative clause, as well as instances in conditional clauses which were also interpreted as non-assertive.

Affirmative

(74) The goal of detecting deception requires far more public scrutiny than it has had up until now. As a society, we **need to** have a very serious conversation about this. (*General prose*, E25)

Negation with an auxiliary do

(75) It was exactly the reaction Jaffe had been hoping for. "I guess I don't **need to** ask whose house we 're looking at." Mohammed knew that he was looking at his nephew 's home and family, but he remained completely impassive. (*Fiction*, N03)

<div style="border:1px solid">Negation with no</div>

(76) Women have "so much," former President Ronald Reagan says, that the White House <u>no longer</u> **needs to** appoint them to higher office. (*General prose*, F42)

<div style="border:1px solid">Hidden negation</div>

(77) I can't even keep a stupid houseplant alive. I mean, how hard is that? <u>All you</u> **need to** <u>do is water the damn thing</u>. (*General prose*, G36)

<div style="border:1px solid">Question</div>

(78) "I 'm almost finished," he said, thinking he might be in the way, though there were other sinks. <u>Why did</u> he **need to** use this one? He noticed a faint metallic odor. (*Fiction*, L21)

<div style="border:1px solid">Comparative clause</div>

(79) It was a close call, but I figured in the scheme of things, my mother needed me to go with her <u>more than I</u> **needed to** stay. (*Fiction*, G36)

<div style="border:1px solid">Conditional clause</div>

(80) "I want to do more. I understand if you don't want to. <u>If</u> you **need to** rest." (*Fiction*, P19)
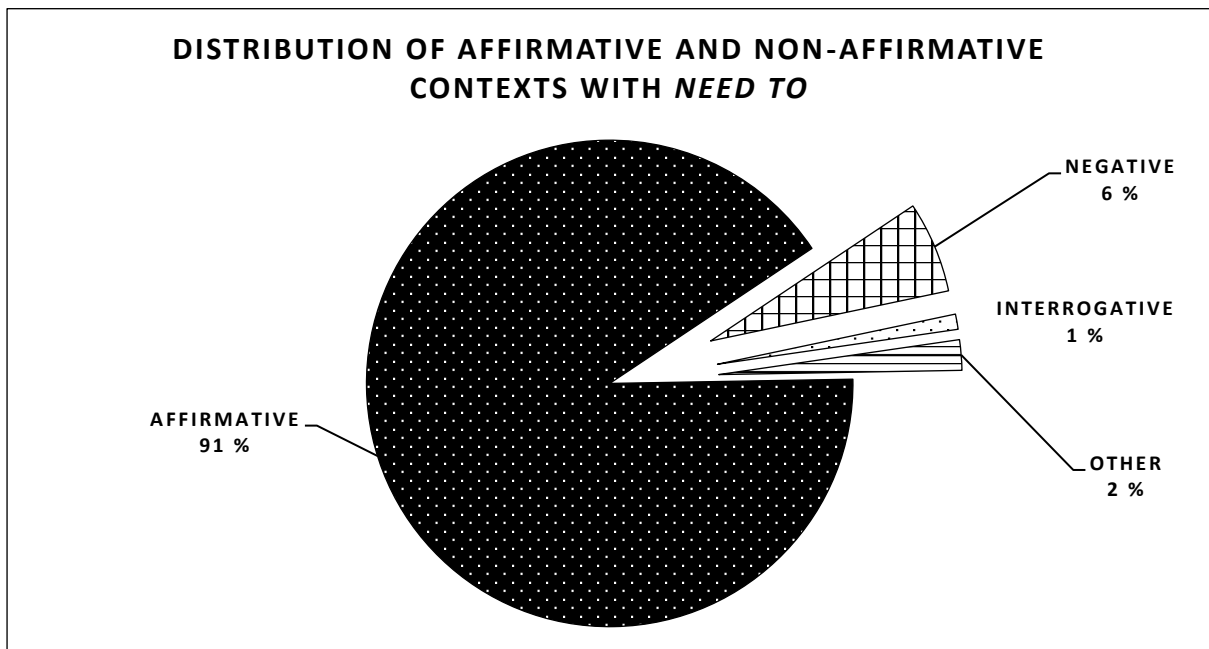
Figure 7. *Distribution of **need to** between different context types in AmE06*

Another feature that distinguishes semi-modals from core modals is that they accept another modal auxiliary as a modifier. Four different modal auxiliaries were found accompanying *need to* in AmE06. The most common auxiliary is *will* with 15 instances. *Would* is used six times, *may* four times and *might* once. Although the total number of modifying modals is low in the data the proportions are like in Nokkonen (2015: 48) who examined the modifiers modifying *need to* in spoken British English. Percentages are shown in figure 8. Since Nokkonen studies spoken data the values are not directly comparable. The LL test shows that the difference between *will* and *would* is not statistically reliable, but the difference between *will* and *may* is.

Modified by *will*

(81) If this trend continues as expected, the Administration and Congress will **need to** consider ways to address Medicare's finances. (*General prose*, H20)

Modified by *would*

(82) Developing countries interpreted the 'common but differentiated' language with great precision: industrialized nations would **need to** take the lead by cutting their own emissions and transferring large sums of environmental assistance to the South. (*Learned*, J08)

Modified by *may*

(83) IM can " jump ports, " so you may **need to** consider multiple measures, such as a firewall appliance, software solutions, or locking down workstations. (*General prose*, F45)
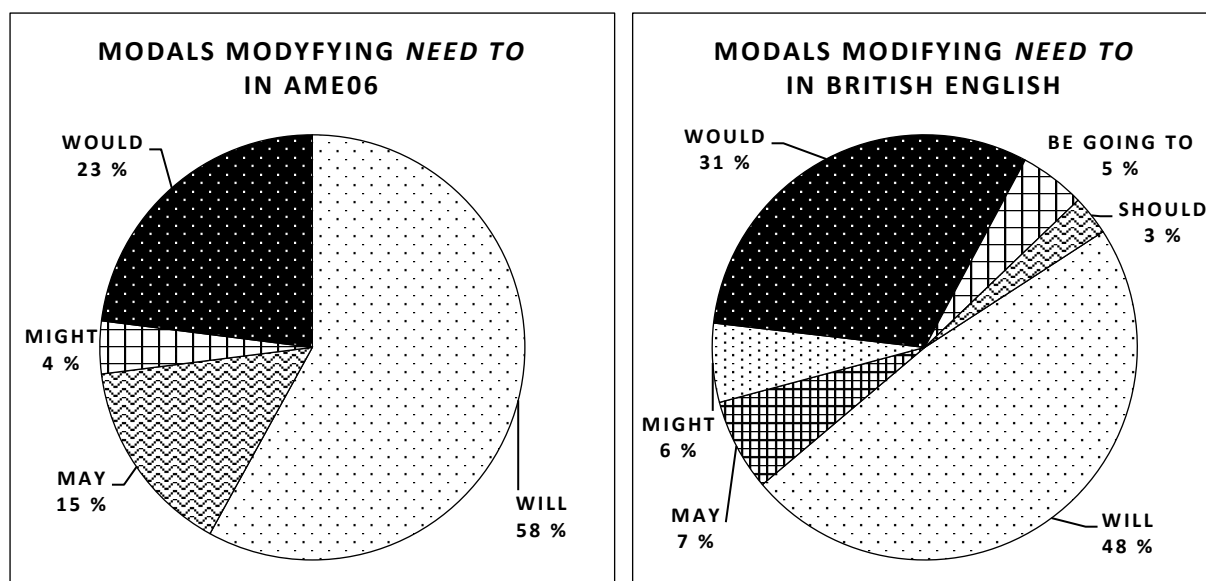
Figure 8. *Modals modifying **need to** in AmE06 and in British English (Nokkonen 2015: 48)*

4.5. Deontic *need to* on the weak-strong cline

In this section the aim is to apply the fuzzy set theory developed by Coates (1983) to the semi-modal *need to* in the AmE06 corpus. This will be done by applying the matrix introduced in section 2.3. in table 4 as extensively as possible to each of the instance and as a result place each instance on the strong-weak continuum. The section is divided into subsections according to the subject type in the following order: first-person singular, first-person plural, second person, third person animate, third person inanimate, and existential *there*. Relevant, illustrative examples are provided from the corpus data in each section which will be discussed in terms of the framework. Furthermore, a table is provided after each corpus text extract where the example in question is plotted against the matrix, which should make the reasoning more explicit to the reader. Each instance is placed on Coates' cline accordingly. The parameters and the matrix were presented in section 2.3., but it was decided to add them here for the sake of the reader.

(a)   second person subject
(b)   subjective
(c)   speaker has authority over subject
(d)   verb is agentive
(e)   paraphrasable by 'it is obligatory/absolutely essential that'
(f)   animate subject
(g)   paraphrasable by 'it is important that'
(h)   inanimate subject

Table 4. *Matrix to indicate the relative strength of deontic/dynamic (root) use of **must*** (Coates, 1983: 37)

a	b	c	d	e	f	g	h
+	+	+	+	+	+	-	-

4.5.1 First-person singular

Instances with a first-person singular subject do not fit well into the framework since the speaker and the subject are essentially the same person. Hence there are no authority structures to be found in first person singular instances. Journalists and academic writers tend to avoid the first person in their writing, thus all the instances with a first-person singular subject are found from the categories of *general prose* and *fiction* in the corpus. In all the first-person instances the subject is obviously animate.

None of the 34 instances land on the core or the periphery. They all resemble the skirt with stronger and weaker examples. In most of the instances the speaker is simply communicating his or her needs, hence most are interpreted as objective rather than subjective. Most instances are placed in skirt 3. Adverbs as well as the context help determine the strength of an instance. (84) and (85) are examples of stronger instances in the group. They are modified by adverbs *desperately* and *clearly*, which makes them more favorable to be paraphrased as 'it is absolutely essential that'. In (84) the main verb is a cognitive process, *believe*, whereas in (85) the main

verb is an action verb, *talk,* which would make the latter a little stronger than the former. In the former the need to perform the action seems to arise more within the speaker than in the latter, although this is difficult to tell.

There was one instance with a first-person subject where there was an explicitly mentioned external speaker. In (86) the speaker is the shrink, who can be interpreted as being in authority in relation to the subject, I. Here it is questionable whether the deontic source is the shrink, who is suggesting the subject to perform the action, or the subject, to whom it would be beneficial to talk to his or her father. It would be unlikely to think that a shrink would strongly oblige their client to do something, and hence this instance would probably be rephrased as 'it is important that'. This instance in nevertheless placed in the Skirt 1 group based on the existing authority structure. (87) is an example of a weak instance where need to functions closely to its lexical meaning. The deontic source here is clearly the subject's internal, physiological need to cry, something that the speaker has no conscious control over. To burst into tears works better as a happen-clause than a do-clause, so here the verb is interpreted as non-agentive. This is the only instance in the corpus where need to is in gerund form. This instance is the weakest from these examples. The total numbers are presented in figure 9.

(84) "Yeah, but she told me she'd be out by Monday." I was relieved when Herma decided to leave it at that. The story sounded a little thin, even to me, but I desperately **needed to** believe it. Then, without even a hint of foreboding, I made my first executive decision in the health care arena." (*General prose*, G53)

a	b	c	d	e	f	g	h
-	+	-	+	+	+	-	-
Skirt 2							

(85) Dr. Skiller was no doubt making sure the latest delivery of fresh vegetables was one hundred percent organic. <u>Clearly</u>, I **needed to talk** to Dr. Zumi and the personal trainer before I tackled the patients and the terrified staff of Mexicans - none of whom spoke English. (*Fiction*, L08)

a	b	c	d	e	f	g	h
-	-	-	+	+	+	-	-
Skirt 2							

(86) …admired and wanted to be like her, so much that they had me seeing a shrink by the time I was three. <u>The shrink said</u> I **needed to** spend more time with my dad. But how? Mom was irresistible. (*Fiction*, R04)

a	b	c	d	e	f	g	h
-	+	+	+	-	+	+	-
Skirt 1							

(87) "Since school," I mumbled, searching for a way to say that, while I was enjoying our chat, I should really get going as I 'd be **needing to** <u>burst into tears</u> soon. The receptionist, more eager to rescue Charlie than me, reminded him of an impending meeting. (*Fiction*, R04)

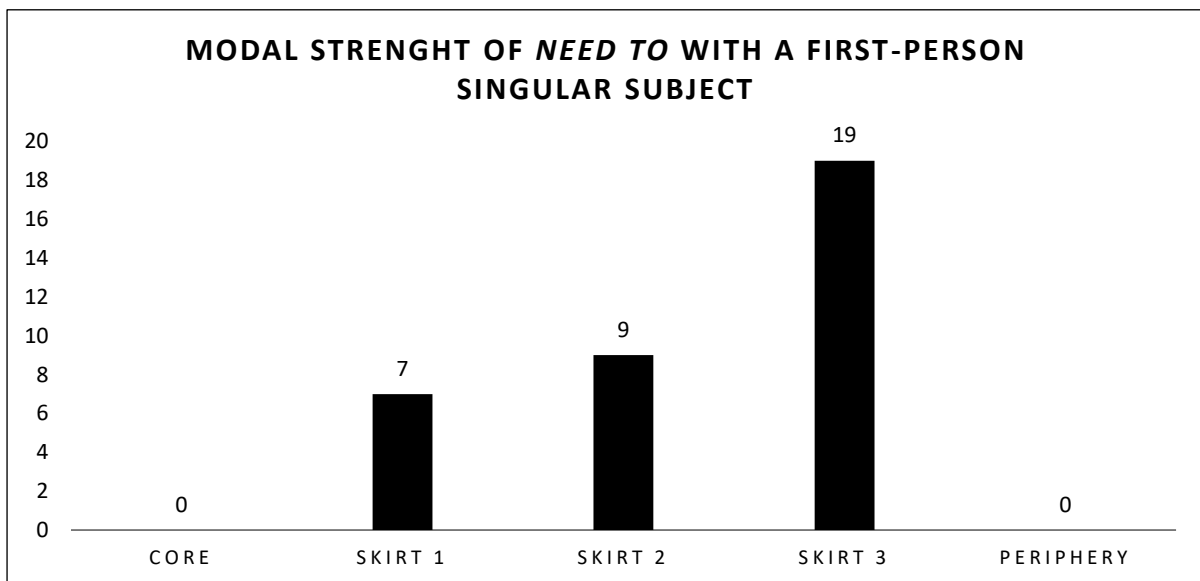| a	b	c	d	e	f	g	h
-	-	-	-	+	+	-	-
Skirt 3							

Figure 9. *Modal strength of **need to** with a first-person singular subject*

4.5.2 First-person plural

Occurrences with a first-person plural subject are mostly weak and objective, similarly to the first-person singular cases. There are cases where the speaker simply communicates the need of the groups he or she is representing, and cases where the speaker tries to influence the implied audience. The latter being stronger than the former. In a few instances it is possible to interpret that there exists an authority structure between the speaker and the recipient, moreover, a sense of urgency or importance may be sensed from the context. All the instances were placed in the skirt, mostly in the third group. The stronger instances were placed in the second group. There were instances from all the four subcorpora.

(88) is an example of a weak, objective case where the speaker communicates the needs of his group to an equal recipient, combined with a non-agentive verb *see*. This is the most common type with first-person plural subjects. (89) is an example of a stronger instance. Here the writer is communicating the urgency that Georgia implement alternative transportation methods, furthermore, the pronoun *we* could also be replaced with *you* and the directive function is further strengthened by the adverbial *right now*. There is also a sense of the speaker's involvement in this instance. This type of use seems to be more typical in the press genre.

In (90) the deontic source is God who can be argued to be an authority. Thus, this is an example of a subjective case where an authority permits the recipient to perform an action. This might be a core case, however, because it can hardly be interpreted that the speaker (God) here is imposing an obligation on the recipient and it would make more sense to rephrase the sentence that "it is important that we do what we need to do to restore the health of our souls", this instance is put into the strongest group in the Skirt.

(88) "Where's Antonio?" "In … in the kitchen," Joe admitted reluctantly." He's helping Mama." "Tell him <u>we</u> **need to** <u>see</u> him." Frank glanced around the noisy room. (*Fiction*, L10)

a	b	c	d	e	f	g	h
-	-	-	-	-	+	+	-
Skirt 3							

(89) …the traffic congestion the region is experiencing will almost certainly worsen. Georgia needs to implement every form of alternative transportation available, and <u>we</u> **need to** do it <u>right now</u>. Commuter rail, including the Atlanta-Lovejoy line and the Atlanta-Athens line… (*Fiction*, L10)

a	b	c	d	e	f	g	h
-	+	-	+	+	+	-	-
Skirt 2							

(90) And don't use it as an excuse for becoming self-absorbed, egotistical or selfish. Use the time wisely by becoming more dependent on <u>God</u>, who gives us permission to do what <u>we</u> **need to** do to restore the health of our souls. (*Press*, C06)

a	b	c	d	e	f	g	h
-	+	+	+	-	+	+	-
Skirt 1							

Overall, the picture is similar with both first-person singular and plural subjects. The instances are weak and objective. Figure 10 shows how the instances with first-person plural subject are distributed on the cline.

MODAL STRENGHT OF *NEED TO* WITH A FIRST-PERSON PLURAL SUBJECT

35

30

25

20

15

10

5

0

| CORE | SKIRT 1 | SKIRT 2 | SKIRT 3 | PERIPHERY |

0   1   4   32   0
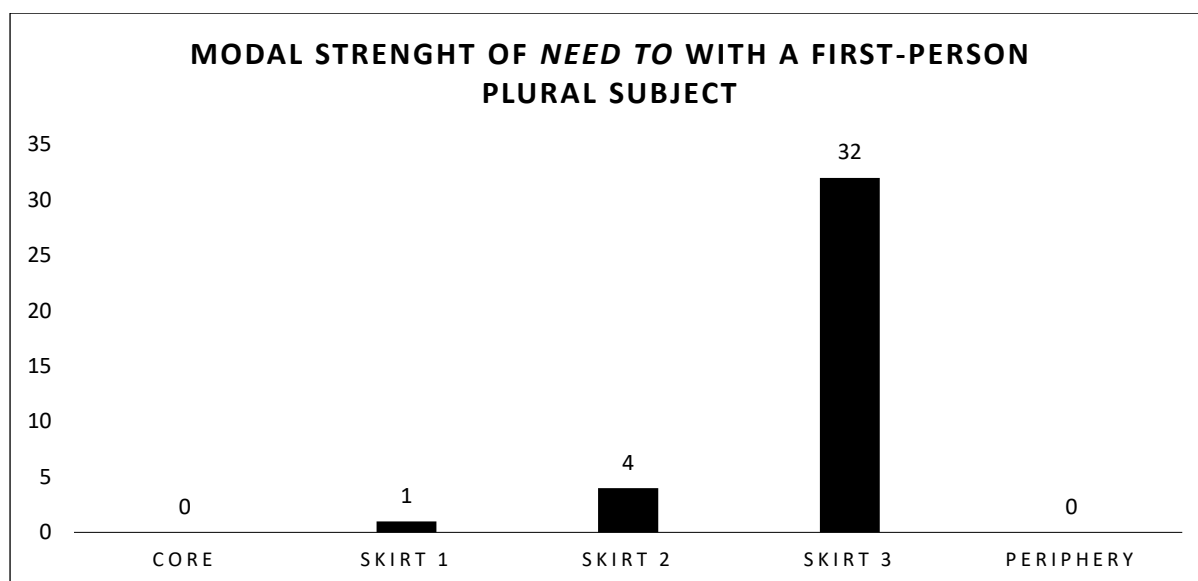
Figure 10. *Modal strength of **need to** with a first-person plural subject.*

4.5.3 Second person

Instances with a second person subject are found mainly in the *fiction* and *general prose* sections, there being one example from the *press* category and five from *learned*. Typically you is used in an impersonal way, there is no person in particular the speaker or, more commonly, thewriter, is addressing. This is a typical use to manuals and instruction books which are found in the *general prose* -subcategory, such instances are weaker compared to those where there is a single articulated addressee. Similarly, four out of five instances in the *learned* category were from a single text source, namely an introduction to the different academic citation styles, in all of which the addressee is the impersonal reader.

The strongest instances with a second person subject were found in the fiction section, since directly addressing another person and giving directives is more natural to spoken than written language, and naturally most text extracts in the corpus that include spoken dialogue come from fiction. The problem, however, with these instances is that it is nearly impossible to interpret

whether there exists an authority structure between the speaker and the recipient without reading the whole book. Neither is there such metadata available for the AmE06 corpus.

The first to examples (91) and (92) illustrate the weaker end of the cline. The first one is the only example from the press category, it is fairly objective, and the speaker is addressing the public in general. The latter one is from the book on citation styles, the instance is a little bit stronger as it seems that the writer is expressing his or her subjective opinion on the matter, although it remains unclear since he or she could as well be referring to the implied external set of conventions of the academic world of writing. The verb is an agentive activity verb, and there is nonetheless a greater sense of importance in this instance. In the following two examples the deontic source is the speaker him/herself and they are directly addressing another person, hence in both the speaker involvement is strong. In the first one there is no authority structure as the speaker and the addressee are siblings, furthermore, the verb *be* is not agentive. The latter example, in contrast, involves an elderly person giving orders to an adolescent, hence she is in authority. The verb *help* is also agentive. Thus (94) is an excellent example of *need to* in its core deontic use.

There were two instances from one text example where *need to* appears with a second person subject (95), but the speaker is addressing himself in an inner dialogue trying to motivate himself into an action. Although the speaker and the addressee are the same person, like is typically the case with first-person subjects, in these examples the speaker is not simply communicating his needs to an audience. Although the verbs here are mental, cognitive processes, the sense of urgency is clear from the context. It is also possible to interpret there existing a two separate levels of consciousness in the same physical person, one who acts as

the rationale (the authority) and the other who is the person in panic. These two were both placed in the core although they are not the clearest examples.

(91) In marketing circles, the trend of holiday shopping starting earlier and earlier is known as "Christmas Creep." "It's a retail mentality that you **need to** be first out of the gate, and people keep making the race longer and longer… (*Press*, A19)

a	b	c	d	e	f	g	h
+	-	-	-	-	+	+	-
Skirt 3							

(92) First, good citations parcel out credit. Some belongs to you for the original work you did; you **need to** take full responsibility for it. Some belongs to others for their words, ideas, data, drawings, or other work. (*Learned*, J75)

a	b	c	d	e	f	g	h
+	+	-	+	+	+	-	-
Skirt 1							

(93) Moreover, I couldn't stop thinking about what my sister had said to me once, as she was breast-feeding her firstborn: "Having a baby is like getting a tattoo on your face. You really **need to** be certain it's what you want before you commit." (*General prose*, G39)

a	b	c	d	e	f	g	h
+	+	-	-	+	+	-	-
Skirt 1							

(94) I 'm saving up for a glow-in-the-dark bead curtain that I saw last week in Spencer Gifts. Bedtime is at eight, Mrs. Johnston says. They don't need baths, but you do **need to** help them with their teeth. (*General prose*, G55)

a	b	c	d	e	f	g	h
+	+	+	+	+	+	-	-
Core							

(95)   I was tingling with fear, and all I wanted to do was to turn and sprint
       for shore. But I told myself, Stay calm. <u>You</u> **need to** focus. <u>You</u> **need
       to** figure out what this is. Taking a deep breath, I looked down into
       the deep black sea. (*General prose*, G41)

a	b	c	d	e	f	g	h
-	+	+	+	+	+	-	-
Core							

Figure 11 shows how the instances with a second person subject land on the continuum. There were a few core instances which fulfilled at least five of the first six parameters. Instances in skirt 1 included a distinguished speaker and addressee, however it was either impossible to say whether there was an authority structure or not, or if the speaker him/herself was the deontic source or if they were for example referring to an objective set of rules or conventions. Most instances were placed in the skirt 2 group. In these instances, *you* is typically used as a rhetorical device such as in a press speech, or as an impersonal subject as in instructions or manuals.
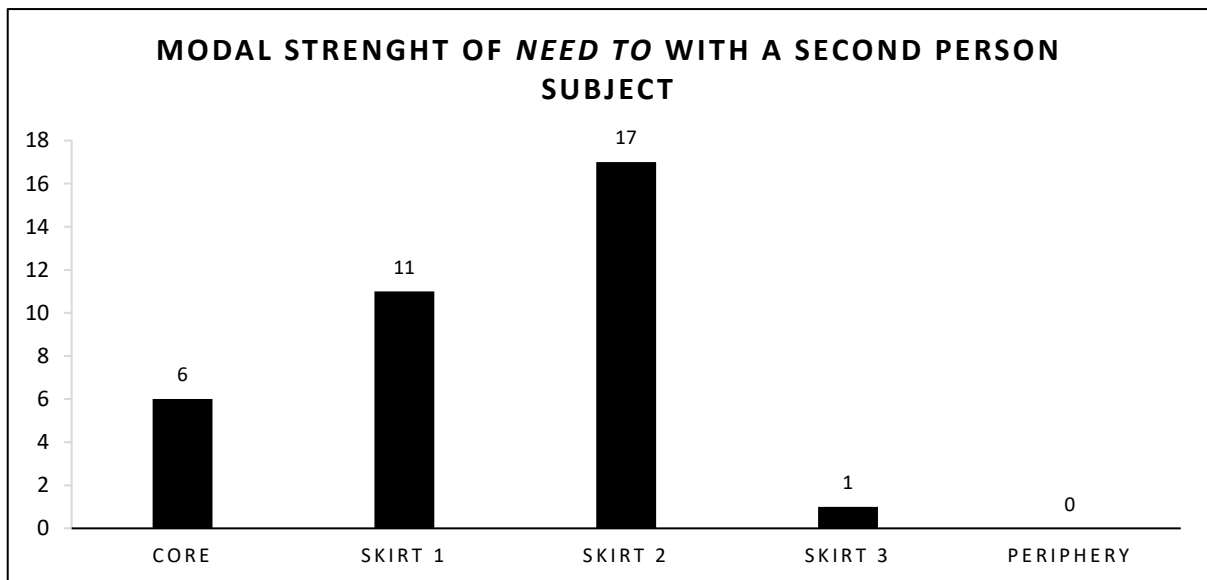

Figure 11. *Modal strength of **need to** with a second person subject*

4.5.4 Third person animate

Third person subjects are found in all the four subcorpora. Instances with a third person subject are generally weaker than those with a first or second person subject (Coates 1983: 37). This is also the case in AmE06, as in most instances the speaker's involvement is minimal. In the press genre the addressee is commonly a collective noun such as "Republicans" or "educators", or even "the world". The same applies to most instances in the fiction and learned -subcategories. Most subjects are a collective, there is no single person named. This instantenously makes most instances weak. In fiction the subject is typically a single named person, however in nearly all the cases the speaker simply communicates the objective needs of said person.

(96) is from the press -subgenre, and includes an extract from an interview where the speaker is in authority and expresses his personal opinion on the matter, close to giving an obligation. This example is close to the core, as the verb is also an agentive activity verb. However since the recipient is an institution "the school district" rather than a distinct person, this instance is placed in the skirt 1.Another strong case is (97), in here the immediate context gives us a sense of *need to* being used as a strong imperative. The speaker is an old man, and although it is not completely clear, it can be argued that age provides authority to the speaker here. The main verb is an existential verb *be*, which makes the instance a little weaker. This instance resembles a core case, but the imperative here is indirect, the speaker is not directly addressing the person he would like stop being so selfish, but is commanding the boy to deliver that message to his father. Thus, this instance was placed outside the core in skirt 1.

(98) illustrates a use of *need to* where the speaker is moderately involved but where the obligation is weak, since the addressee is a group and not a single person. These are common in the press and learned -subcorpora, and were placed in skirt 2. (99) is an example of a typical

case found in fiction, where the speaker is communicating the needs of the addressee. These

are placed in skirt 3.

(96) Westchase Community Association President Daryl Manning said the school district **needs to** provide Westchase with proof that the athletes are insured and that the coaches are certified lifeguards. (*Press*, A24)

a	b	c	d	e	f	g	h
-	+	+	+	-	+	+	-
Skirt 1							

(97) The old man nodded, and his voice quivered with frustration and anger. "You tell your daddy that he **needs to** stop being so goddamn selfish and think about the rest of the people in this county. I 'm not the only one that says so. A lot of people around here are fed up. You tell him that from me." (*Fiction*, K07)

a	b	c	d	e	f	g	h
-	+	+	+	-	+	+	-
Skirt 1							

(98) At UGA, the number of men in freshman classes has dropped from 43 percent in 1998 to 37 percent in 2006. Educators **need to** do a better job preparing boys for college, and once they are there, do a better job of encouraging talented male students to seek advanced degrees as a "valuable alternative" to entering the workforce right out of college, Adams said. (*Press*, A19)

a	b	c	d	e	f	g	h
-	+	+	+	-	+	+	-
Skirt 2							

(99) Thinking about Fiona, Gray was reminded that he **needed to** check on Ryan. (*Fiction*, N16)

a	b	c	d	e	f	g	h
-	+	+	+	-	+	+	-
Skirt 3							

The overall distribution is shown in figure 12. No cases were placed in the core or in the periphery. Most cases were in skirt 3 with a couple strong cases in skirt 1 and skirt 2. Compared to the instances with a first-person singular subject, the two charts look quite alike.
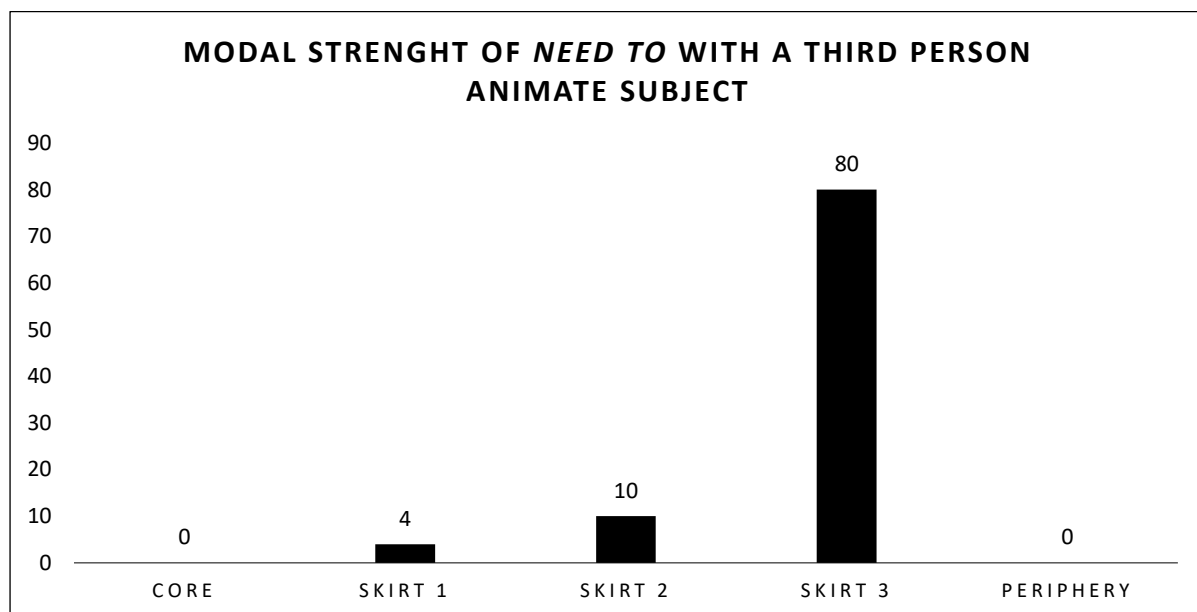
Figure 12. *Modal strength of **need to** with a third person animate subject*

4.5.5 Third person inanimate

There was a total of 16 instances in the corpus with a third person inanimate subject. Deontic meaning of *need to* with an inanimate subject tends to be weak since an inanimate subject cannot directly be commanded to perform an action. A typical instance describes the necessary qualities of the inanimate entity in a certain situation. (100) exemplifies such an instance. Here the writer is describing a medical procedure, and although the verb describes an activity, this is a case in the periphery. (101) is another example of a similar objective and weak instance from an academic text.

There were couple instances that allowed a more subjective interpretation and land in the skirt rather than the periphery. In (102) the sentence allows a subjective interpretation; it is possible that the speaker is expressing his opinion on the importance of preserving the building. Although the subject is inanimate, there is an external addressee or audience whom the speaker can try to influence. The adverb *really* emphasizes the deontic function of the clause. This instance is, nonetheless, unclear since it might as well be that the speaker is referring to some external conditions based on which it is likely that the building is preserved, and hence this instance would be in the periphery and also allow an epistemic reading. Another case placed in the skirt rather than the periphery is (103) in which there is an external addressee (impersonal *you*) outside the inanimate subject (*your firewall*), whom the writer can influence. This instance is otherwise weak, the verb is stative, and the writer is not in authority.

(100) An injection of contrast dye lets Cho and Getzen see exactly where blood flow from the kidney joins the vena cava. This is the precise point where <u>the filter's tip</u> **needs to** go. Cho carefully slides the sheathed filter along the guide wire and up the vein, stopping after about 10 seconds. (*General prose*, E04)

a	b	c	d	e	f	g	h
-	-	-	+	-	+	+	+
Periphery							

(101) We note also that this segment is in syllable-initial position, which coincides with word-initial position, and that the entire word is dominated by a single syllable. <u>The gesture-calculations component</u> **needs to** turn this representation into a series of instructions to the musculature of the vocal apparatus which produces the relevant acoustic event. (*Learned*, J34)

a	b	c	d	e	f	g	h
-	-	-	+	-	+	+	+
Periphery							

(102) "If they want to move the library and can get a good library, as a citizen, I don't want to pay through the nose," he says. "The Mies building really does **need to** stay. I hope they get a good library." (*General prose*, E17)

a	b	c	d	e	f	g	h
-	+	-	+	-	+	+	+
Skirt 3							

(103) If so, be aware that a variety of new threats are specific to this technology, and your firewall **needs to** be robust enough to handle them. Be sure that the company you choose to install your VoIP products will provide you with specific security advice and protection so that outside IP-specific attacks cannot get through. (*General prose*, F45)

a	b	c	d	e	f	g	h
-	+	-	+	-	+	+	+
Skirt 3							

As can be seen from figure 13, all instances with a third person inanimate subject were weak, with 15 falling into the periphery and 2 in the skirt.
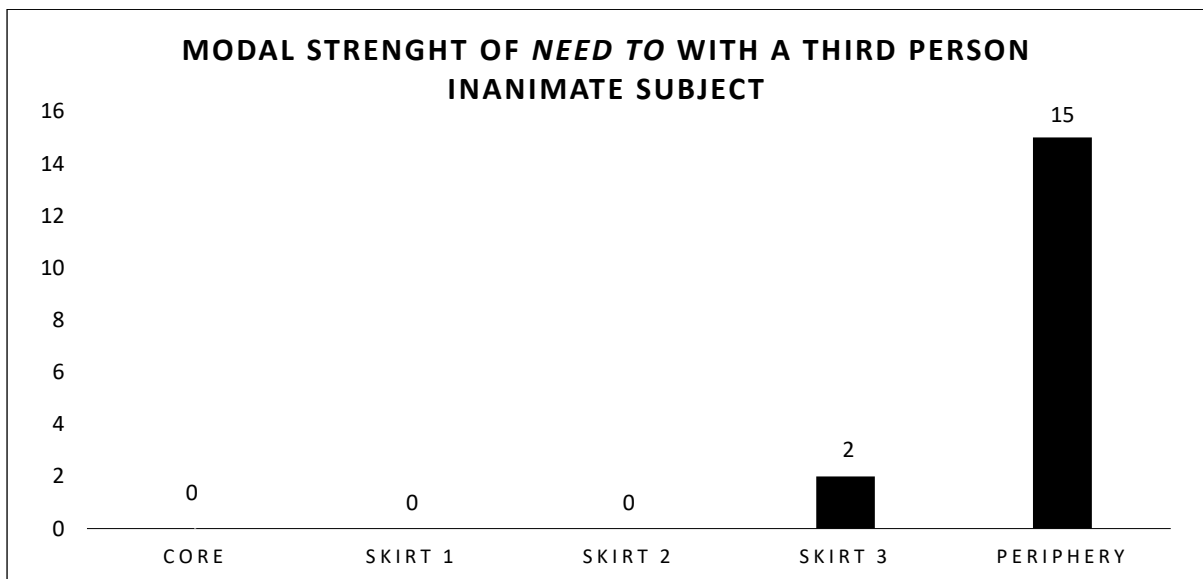

Figure 13. *Modal strength of **need to** with a third person inanimate subject*

4.5.6 Existential there

There was one instance in the corpus where the subject was an existential *there*. Existential *there* is associated with epistemic meaning (Nokkonen 2015: 102), however, in this extract *need to* seems to indicate a deontic function. The speaker uses *ought to* in the preceding sentence, also with an existential subject, where he stresses the importance that there should be a cap on how the executives are rewarded. The two auxiliaries can be seen to work in parallel. Thus, the epistemic sense can be rejected here. It would make more sense to rephrase the latter sentence that "it would be important that there were massive public and private efforts to retrain workers for new jobs", as the speaker is influencing the addressee rather than pondering whether or not such efforts have been done. This instance is still deontic while it is clearly in the periphery.

(104) Before this recourse is taken, in my view, there ought to be a limit on golden parachutes and a cap on the salaries of corporate leaders, particularly when the survival of an industry is endangered. Moreover, <u>there</u> **need to** be massive private and public efforts to retrain workers for new jobs. (*General prose*, G25)

a	b	c	d	e	f	g	h
-	-	-	-	-	-	+	+
Periphery							

4.5.7 Passive

Passive constructions with need to appear in all the four subcorpora. In the English short passive construction the agent is not explicitly stated. Thus the analysis of passive instances in the corpora becomes complicated as there is no agent the speaker can influence. There is one instance where the agent is stated with a *by*-phrase, and few where it is possible to interpret the agent from the context. Instances from the fiction section include written dialogue, and it is

easier to see who the addressee is the speaker may try to influence, but mostly passive constructions appear in formal, written language, and the addressee is ambiguous.

(105) is from a horoscope magazine and the only long passive construction in the corpus where the agent is explicitly stated. This instance is nonetheless weak, as the writer is communicating the conditions that apply in the situation, the work of man can only be finished by a higher power. Moreover, the agent is not a particular entity, but any high power. The subject is animate, but it can hardly be said that the speaker could be in authority here. This instance is placed in Skirt 3. In (106) the implicit subject is *we.* The subject clearly wants himself and the addressees to remember the angels in their lives. Hence this is a subjective instance. The main verb is a cognitive verb rather than an action verb, moreover, the speaker is not in authority of either himself or the audience. This instance is put in skirt 2.

(107) and (108) illustrate the weak end of the cline. Both are objective; in the first one, from an academic text, the speaker refers to the socio-economical circumstances in the society that calls for the need to revise the popularized saying. There is no person in particular the speaker is addressing, and the modifying adverb *might* further mitigates the modal strength of *need to.* The latter example is a typical instance from a instruction text where the speaker is simply communicating the objective needs of an inanimate instrument for it to function properly. The speaker's involvement here is non-existent. The former example is placed in skirt 3, as the sentence with the modal appears in an argumentative context. The latter illustrates the periphery.

The last example (109) was the strongest passive instance found in the corpus. This example is from a fictional text and includes a fictional dialogue. The context makes it clear that there are

two participants in the conversation, although their relationship remains unclear. There is a speaker and an addressee, there is a sense of urgency and importance in the situation. It is absolutely important that they are punished. The verb punish in itself is also a strong action verb that requires a willful agent. The speaker makes his reasoning explicit in the extract, nevertheless he is the deontic source. The passive structure is used to avoid directly commanding the addressee, if the speaker used active voice this would be a core instance. Because it is unclear whether the speaker has authority over the subject or not, and the use of passive, this instance is placed near the core in skirt 1.

(105) The wise founders knew that, to bring this destiny to fulfillment, the work of man would **need to** be completed <u>by a higher power</u>. (*General prose*, F03)

a	b	c	d	e	f	g	h
-	-	-	+	-	+	+	-
Skirt 3							

(106) "Had the Blooms not needed a bathroom on Newbury Street that afternoon, I would not be talking about angels today," he says. "For who accomplishes anything without the angels in our life? They **need to** be remembered, if not always in words, then always in our hearts. I do." (*General prose*, E34)

a	b	c	d	e	f	g	h
-	+	-	+	-	-	+	+
Skirt 2							

(107) Popularized by President John F. Kennedy, it generally refers to how a growing economy benefits everyone. These days, however, it might **need to** be revised to say: "A rising tide lifts all yachts." Or perhaps it should be retired entirely, because it no longer appears to be accurate. (*Learned*, J23)

a	b	c	d	e	f	g	h
-	-	?	+	?	-	-	?
Skirt 3							

(108) Some multis tack well, while others won't go through the wind in any kind of seaway and **need to** be gybed instead. In addition, both boat speed and the point of sail can represent significant variables that make it difficult to recommend a standardized procedure. (*General prose*, E06)

a	b	c	d	e	f	g	h
-	-	?	+	?	-	+	?
Periphery							

(109) She studied her chopsticks, then looked me in the eye and asked, "What will it cost?" "More than you can offer. Important people have done bad and dishonorable things. They deserve to be punished. They **need to** be punished." (*General prose*, E06)

a	b	c	d	e	f	g	h
-	+	?	+	+	+	-	-
Skirt 1							

The overall picture in passive voice falls between third person animate and inanimate subjects. Illustrated in figure 14.
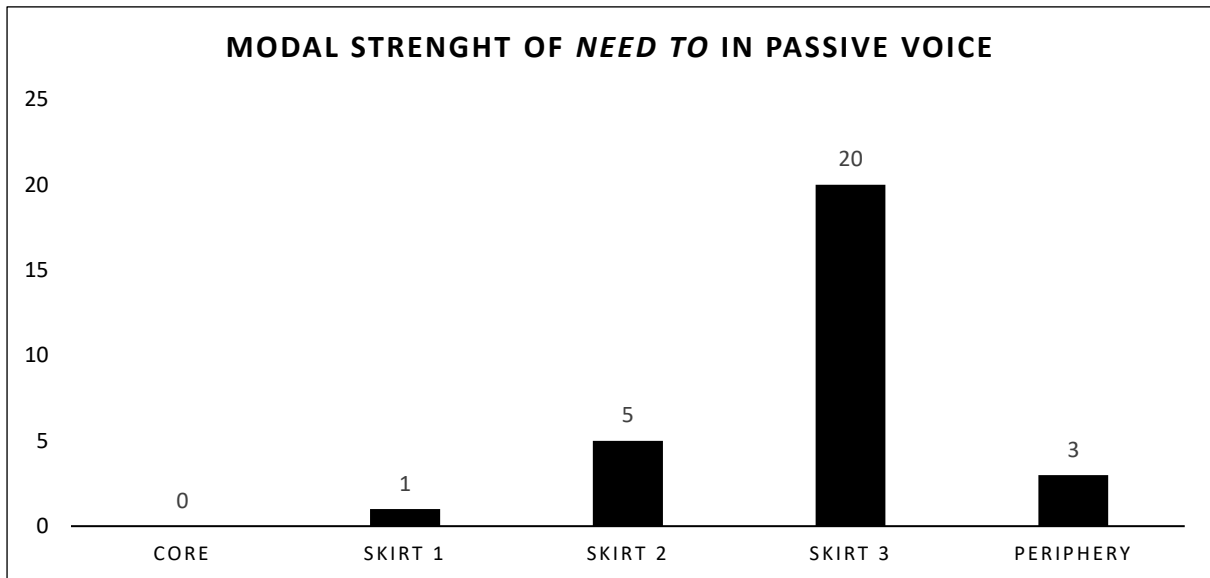

Figure 14. *Modal strength of **need to** in passive voice*

4.5.8 Summary

Figure 15 presents the final number of instances of *need to* in each category on the weak-strong cline. Majority of the instances fall into skirt 3. This is in line with the idea that *need to* is a semi-modal that normally communicates weak obligation and necessity. The differences between the groups are statistically significant according to the LL test, with the exception of periphery and skirt 1.
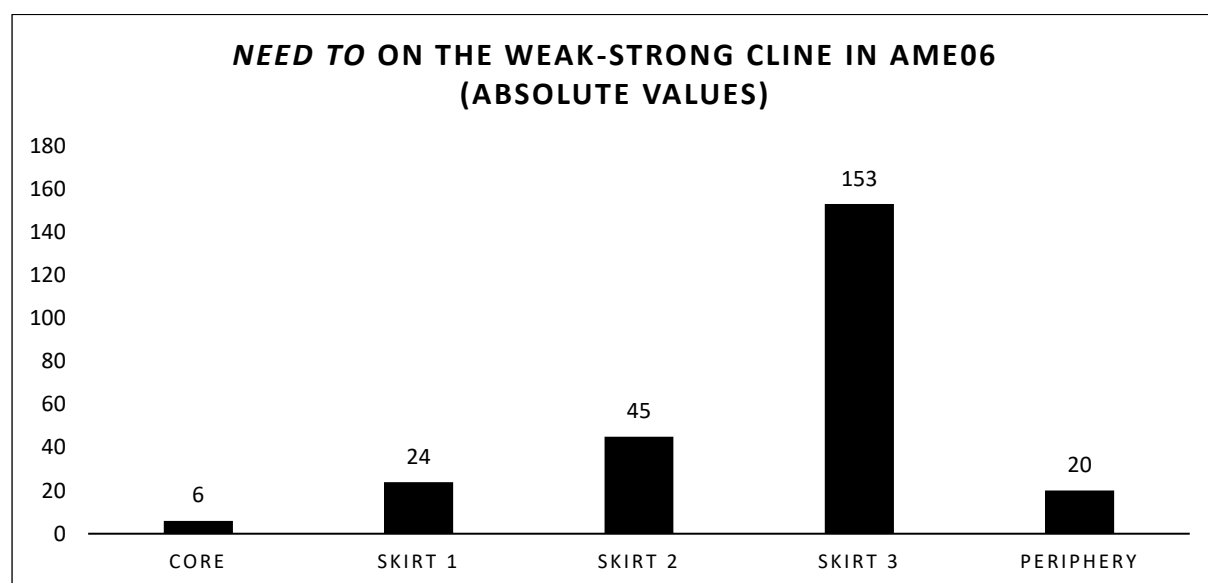

Figure 15. ***Need to*** *on the weak-strong cline in AmE06*

Nokkonen (2015: 106) conducted a similar analysis on written English data using the FLOB corpus. The normalized frequencies are compared in figure 16. The crude picture looks different when looking at the graph. The LL values between the two corpora are as follows, core (7.38), skirt 1 (7.31), skirt 2 (22.73), skirt 3 (16.76), and periphery (0.0), so the differences are statistically significant in all the groups besides periphery. The subjective nature of the analysis plays in important role here, e.g., whereas Nokkonen did not count any of the instances in her written data as core examples, six instances were found in AmE06. An important factor here is the projected authority of the speaker. There is no metadata available for AmE06 that

would explicitly state the relationship between the speaker and the addressee, the existence of an authority structure can only be inferred from the text extract itself, which is a challenging task. Nokkonen comments that she placed most instances with a passive subject to skirt 2, while in AmE06 most ended up in skirt 3. This is perhaps one of the reasons for the difference in the figures.

Nevertheless, there are more instances in the core and skirt 1 groups in AmE06 than in FLOB, which would indicate that *need to* has started to lean more towards the meaning of the core modal *must*, which communicates strong obligation and necessity. The frequency differences in the skirt 2 and 3 groups are major and looking at these two groups FLOB would seem to have stronger instances than AmE06. However, in many instances in AmE06 the decision between skirt 2 and skirt 3 would prove to be difficult and sometimes based on arbitrary factors despite the utilization of the matrix.
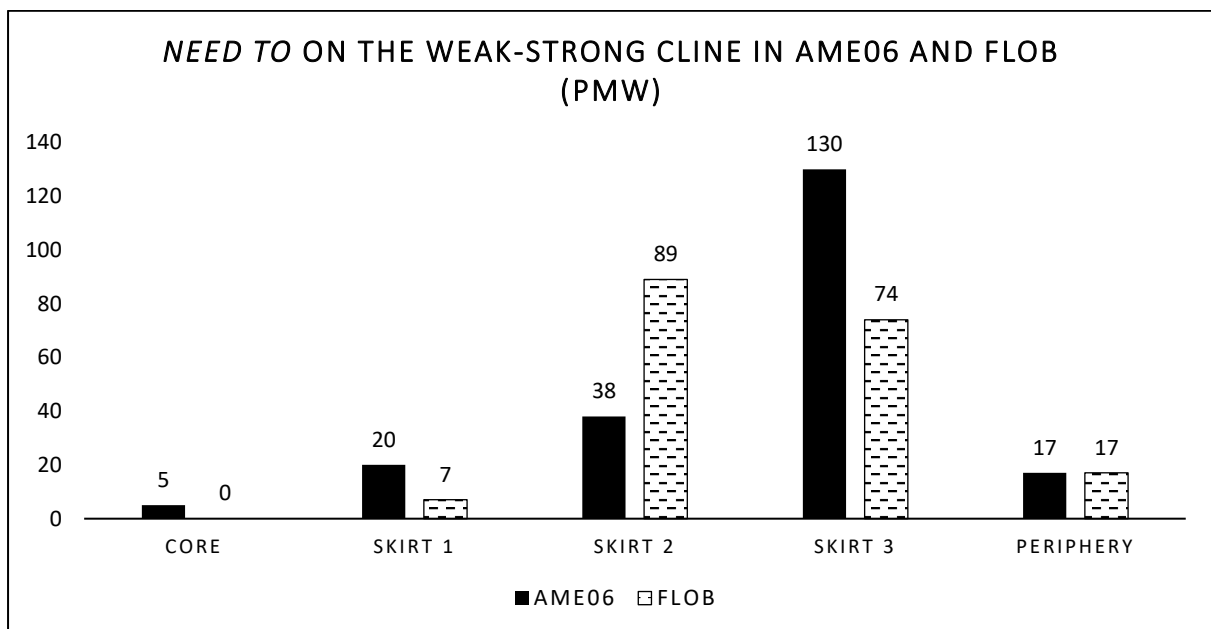

Figure 16. *Need to on the weak-strong cline in AmE06 and FLOB*

4.6. Epistemic uses of *need to*

Epistemic use of *need to* is still a rare, evolving grammatical feature in modern English. Nokkonen, for example, found only two instances from FLOB (2015: 103), while Collins (2009: 73) was able to find 11 instances from the Great Britain section of the International corpus of English. Epistemic meanings of need to develop from dynamic meanings, as illustrated in figure 2, and hence they share many of the same characteristics. The subject is normally inanimate and existential, the main verb is stative, and the aspect is either perfective or progressive. Epistemic instances, or instances that tentatively suggest an epistemic reading, would thus most likely be found among the 20 instances in the periphery, which include the instances with an inanimate subject or an existential *there*.

There were, unfortunately, no clear cases found in the AmE06 where *need to* would have been used in epistemic sense. Some instances in the periphery suggest an epistemic reading, but on a closer examination they are all deontic. Two examples are included here. In (110) the main verb is a stative *be* and the subject is an existential there, however, it is in present tense and the speaker is more likely trying to influence the audience and not expressing his confidence that these actions will be taken into action. Hence still a deontic instance. In the latter instance the subject is inanimate, the tense is present, the verb reflect is active and dsecribes an action. There is, nevertheless, no explicit logical deduction present in the context and the deontic interpretation makes sense, although the modifying auxiliary *may* weakens its strength.

(110) Before this recourse is taken, in my view, there ought to be a limit on golden parachutes and a cap on the salaries of corporate leaders, particularly when the survival of an industry is endangered. Moreover, there **need to** be massive private and public efforts to retrain workers for new jobs. (*General prose*, G25)

a	b	c	d	e	f	g	h
-	-	-	-	-	-	+	+
Periphery (Deontic)							

(111) The American Council on Education (2000) raised questions regarding not only the creation of the course but also the revision of the course and the amount of time needed to launch the course. In addition to courseload reductions, workload reductions from out-of-class responsibilities may also **need to** accurately reflect the additional amount of time involved with course creation and modification. (*General prose*, F30)

a	b	c	d	e	f	g	h
-	-	-	+	-	-	+	+
Periphery (Deontic)							

4.7. The semantic domains of *need to*

The instances can be placed into the semantic categories presented in section 2.3. according to their placement on the cline. According to Nokkonen (2015: 109, 153–155) the internal domain should include instances that were placed in skirt 3, core instances and skirt 1 would comprise deontic 1, skirt 2 comprises deontic 2, and finally instances in the periphery form the dynamic meaning. As discussed earlier, there were no epistemic instances found from the corpus. Figure 16 illustrates the distribution between the meanings. A few examples are provided and briefly discussed. The distribution of the instances in the categories is presented in figure 17.
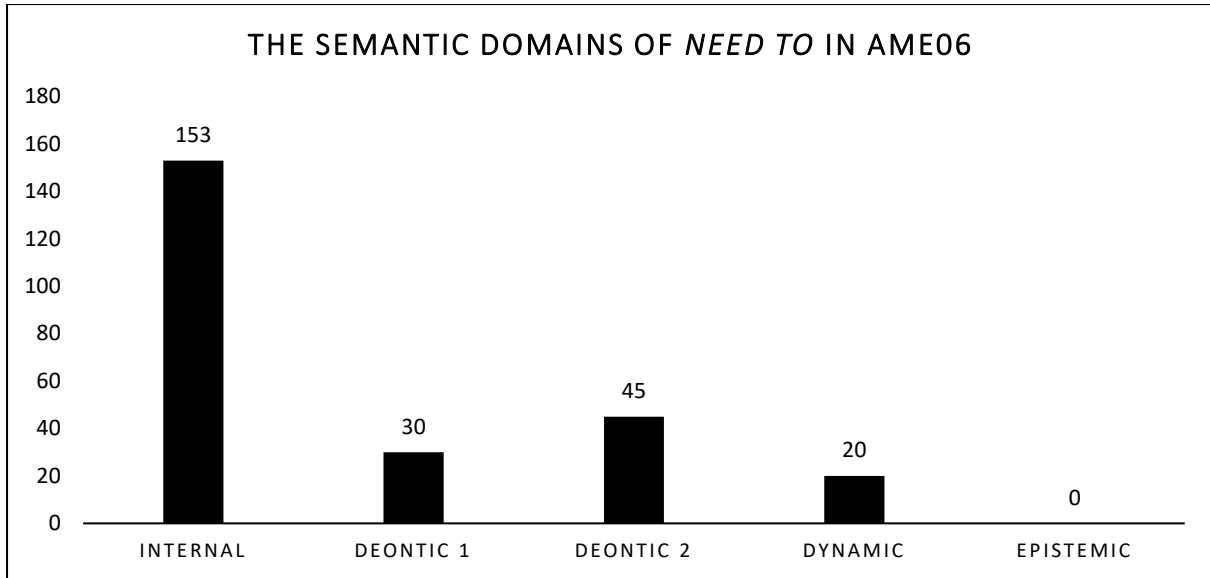
Figure 17. *The semantic domains of **need to** in AmE06*

The internal domain includes instances that semantically communicate basic meaning of *need to*, *i.e.,* internal compulsion. The first two examples clearly both denote the internal, physiological needs of the subject, and hence the necessity arises from the subject itself. The last example is a little more ambigous but still considered internal. There is some overlap between the internal and external factors, as from the context it becomes clear that there is the author's (and other persons') opinion that there should exist personal and ecclestial connection between the recipients. Simultaneously, nonetheless, the need arises from within the subject as interchurch couples need to relate to each other for the sake of their own wellbeing, hence from a semantic point of view this instance is still a member of the internal domain.

Internal

(112) Perhaps the effort was too exhausting, and they **needed to** save their energy for flirtatious glances and fanning themselves. (*Fiction*, P07)

a	b	c	d	e	f	g	h
-	-	-	+	-	+	+	-

Skirt 3

Internal

(113) "Since school," I mumbled, searching for a way to say that, while I was enjoying our chat, I should really get going as I 'd be **needing to** burst into tears soon. The receptionist, more eager to rescue Charlie than me, reminded him of an impending meeting. (*Fiction*, R04)

a	b	c	d	e	f	g	h
-	-	-	-	+	+	-	-
Skirt 3							

Internal

(114) Several authors mention the difficulties of raising children in an interdenominational home. Still others address how interchurch couples **need to** relate both personally and ecclesiastically. (*General prose*, D08)

a	b	c	d	e	f	g	h
-	-	-	+	+	+	-	-
Skirt 3							

The first category in the external domain is deontic 1, which includes all the instances from the core and skirt 1 groups. In the first example the author answers to a question sent in by a reader, and the source of obligation (the author) is external to the subject. The author is a specialist and the addressee a child, so there is a clear authority structure. The second and third examples similarly include an external speaker, who directly addresses another person and almost as if command them to perform an action. However, there is no apparent authority structure and the main verb in either does not describe an action. It seems clear that the deontic source is the speaker in both these examples.

Deontic 1

(115) A. Your son (not you) **needs to** decide whether this class is a good fit. (*General prose*, F11)

a	b	c	d	e	f	g	h
-	+	+	+	+	+	-	-
Core							

Deontic 1

(116) You aren't a student; you're a teacher. You **need to** know how to show others. You had to learn how to do yard duty when you were a rookie teacher. (*General prose*, N18)

a	b	c	d	e	f	g	h
+	+	-	+	+	+	-	-
Skirt 1							

Deontic 1

(117) I warned you when you came here that you **needed to** be careful of him. (*Fiction*, P15)

a	b	c	d	e	f	g	h
+	+	-	+	+	+	-	-
Skirt 1							

In the deontic 2 domain are the instances in which the deontic source is not necessarily the speaker since there are other external factors present. Typically, in these instances the speaker tries to influence people in general, the subject may be an impersonal *you* or a general *we*. For example, in the first instance (118) the speaker is addressing a general audience. Although the subject is in first person, the speaker is clearly not communicating any internal needs. On the other hand, it would be difficult to see how the speaker could impose an obligation to him/herself. The source of obligation in here remains vague. The other two instances are from the general prose category, and in both the author is addressing the impersonal reader with a

second person subject. The author cannot directly command the reader, and either are the subject's internal needs present.

Deontic 2

(118) I think it's taken us several years to get into this situation and it's going to take us several years to get out of it, "Adams said." We may **need to** do some more marketing... (and) we've got to change some social norms. (*Press*, A19)

a	b	c	d	e	f	g	h
-	-	-	+	+	+	-	-
Skirt 2							

Deontic 2

(119) With multihulls there are so many design variables that you **need to** experiment to find what works best with your boat. (*General prose*, E19)

a	b	c	d	e	f	g	h
+	-	-	+	+	+	-	-
Skirt 2							

Deontic 2

(120) Most people don't. The DNA sciences will dominate in the twenty-first century, and you **need to** understand the terms and the concepts if you 're going to stay on top of and benefit from the huge DNA-related advances in medicine and other sciences. (*General prose*, J13)

a	b	c	d	e	f	g	h
+	-	-	-	+	+	-	-
Skirt 2							

Lastly there are dynamic instances that express vague external necessity, e.g. a necessity that arises from the required qualities of the subject referent in a certain situation. The subject is commonly inanimate. Per definition dynamic modality is related to the abilities and volition of

the subject. The need in these examples simply communicates the necessary objective properties of the subject.

Dynamic

(121) If so, be aware that a variety of new threats are specific to this technology, and your firewall **needs to** be robust enough to handle them. (*General prose,* F45)

a	b	c	d	e	f	g	h
-	-	-	-	+	-	-	+
Periphery							

Dynamic

(122) There's a new bolt to use, and any paint **needs to** be scraped away to ensure a good ground. (*General prose,* E16)

a	b	c	d	e	f	g	h
-	-	-	-	-	-	+	+
Periphery							

Dynamic

(123) The gesture-calculations component **needs to** turn this representation into a series of instructions to the musculature of the vocal apparatus which produces the relevant acoustic event. (*Learned,* J34)

a	b	c	d	e	f	g	h
-	-	-	-	+	-	-	+
Periphery							

5. Discussion

This thesis has explored the semantic and pragmatic variation of the semi-modal auxiliary verb need to in contemporary American English utilizing corpus-based methods. The study has aimed to connect the different ways need to is used to the theory of modality and its subgenres deontic and epistemic modality, moreover, the analysis relies on the fuzzy set theory developed by Coates (1983) and the semantic categorization developed by Nokkonen (2015). The 2006

Corpus of American English, which was used as source in this thesis, includes a total of 248 instances of *need to*. All these instances were analyzed manually considering the immediate context that was accesible via the electronic corpus tools, and the possible metadata, i.e., the text type and the nature of the original source text.

The results of the analysis have been compared to previous results from other literature, whenever possible. There have been no similarly conducted comprehensible analyses on need to before Nokkonen (2015), who used primarily spoken British English as data. Considering the initial frequency of need to in the AmE06 corpus, the results are well in line with the other studies, e.g. Leech (2003) and Daugs (2017). The frequency of need to is on the rise in American English, as there were more instances in the AmE06 than there were in the Frown coprus, which spans the years 1991–1992, and less instances than in COCA, when looking at more recent years.

Need to appears in the corpus data with every possible subject type, the most common group being the third person animate. The statistical difference between the other subject types turned out to be insignificat, with the exception that non-intentional subject types (existential *there* and third person inanimate subject) were the least frequent. Nokkonen had analyzed the frequencies of the subject types in the FLOB corpus, and the crude picture is similar, although there are significantly more third person animate instances in the AmE06 than in FLOB and more passive instances in FLOB. It is hard to say where the difference comes from, as Nokkonen does not provide exact numbers for each subject type in each subcategory. *Need to* occurs more frequently in the *fiction* and *press* subcorpora in AmE06 in comparison to FLOB, in which the *learned* and *general prose* are dominant. Third person subject is a typical feature in fiction whereas passive form is common in acamedic texts, this most likely explains the

difference in the frequencies. *Need to* appears primarily in affirmative contexts in AmE06, although there were a few instances in non-affirmative contexts as well. The most common modal *modifying need to* in the data was the core modal will. There were also a few instances with the modals *would*, *may*, and *might*.

All the instances were analyzed and placed on a weak-strong cline, based on Coates' theory (1983), in a similar manner to Nokkonen (2015). Need to is clearly used mostly in a weak deontic sense in the AmE06. This means that in most instances need to is used to communicate objective, internal needs of either the addressee or the speaker. Most instances with a first-person, third person animate and passive subject types fell into this category. Core instances which denoted the strongest deontic use were rare and occurred only with second person subjects. Most cases with a third person inanimate subject fell into the periphery. The final figures were compared to those counted by Nokkonen (2015) who provided numbers from the FLOB corpus.

The overall figures were noticeably different from FLOB. In all the categories except *periphery* the difference was statistically significant. There were no core instances in FLOB, whereas there were 5 instances in AmE06. All the core instances were found from the fiction subcorpus, where they typically occur within fictional, written dialogue. One of the parameters that was used to evaluate the strenght of an instance was the existence of an authority structure between the speaker and the addressee. There was no such metadata available in AmE06 that would have explicitly stated the roles and the relationship between the participants in the fictional dialogue, and hence it was often a result of an ambigous interpretation when deciding whether there was an authority structure present. Hence the difference is more likely a result of a

different interpretation of the context of the instances rather than difference actual differences in the ways *need to* is used in the corpora.

There were more more weak (skirt 3) deontic instances in AmE06 than in FLOB and more stronger deontic (skirt 2) instances in FLOB, so it would seem that the extremes are more prolific in the former. Nokkonen states (2015: 107) that she placed most instances with a passive subject type in skirt 2, however in AmE06 most were placed in skirt 3 as they seemed to denote the weaker use according to the parameters. Overall the results show that need to covers a wide range of different levels of modality, from those that impose a strong obligation on the addressee to those which communicate the mere necessities of an inanimate subject in a given situation.

This thesis has aimed to examine the semi-modal *need to* and its semantic functions in an American context. The results provide an overall picture on how the semi-modal is used in written contemporary American English in the context of modality. There had been no similar study conducted on need to in American context prior to this text. The study has its shortcomings, first of all the definition of modality proved to be a challenging task as definitions in literature are often complex and the terminology is often used in different meanings. Secondly, the analysis of the data is highly subjective, and often based on the author's estimations. Moreover, in the reference literature scholars rarely explicitly state the reasoning behind the analysis of actual text examples. Hence comparing the numbers in this study to other studies has to be treated with caution as there is likely a significant difference in the way the actual text examples are analyzed.

Need to is an emerging expression and it seems clear that it is gaining new ground in the field of modal expressions. Thus studying need to will remain an interesting topic in the future as well. The corpus used in this study was a relativile small one, and there are more extensive and larger copora on American English available that could prove to be useful in studying *need to* in its epistemic meaning. The semantic categorization applied here could also be developed further, and perhaps applied to other modal expressions in English.

6. References

Aarts, B. 2011. *Oxford Modern English Grammar.* New York: Ofxord University Press.

Aarts, B. & C.F. Meyer. 1995. *The verb in Contemporary English: Theory and Description*. Cambridge: Cambridge University Press.

Biber, D., S. Conrad and R. Reppen. 1998. *Corpus Linguistics: Investigating Language, Structure and Use*. Cambridge: Cambridge University Press.

Biber, D. & R. Quirk. 1999. *Longman Grammar of Spoken and Written English.* London: Longman.

Biber, D. & R. Reppen. 2015. *The Cambridge Handbook of English Corpus Linguistic*. Cambridge: Cambridge University Press.

Bybee, J. L. 1988. 'The diachronic dimension in explanation', In: *Explaining Language Universals*. Oxford: Blackwell, 350–379.

Bybee, J. L. & W. Pagliuca, 1985. Cross-linguistic comparison and the development of grammatical meaning. In: Fisiak, J. (eds), *Historical semantics. Historical word-formation.* Berlin: Mouton, 59–83.

Coates, J. 1983. *The Semantics of the Modal Auxiliaries*. London: Croom Helm.

Coates, J. 1995. 'The expression of Root and Epistemic Possibility in English', In: Aarts, B. and C.F. Meyer (eds), *The Verb in Contemporary English*. Cambridge: Cambridge University Press, 141–158.

Collins, P. 2009. *Modals and Quasi-modals in English.* Amsterdam: Editions Rodopi B.V.

Collins, P. 2015. *Grammatical Change in English World-Wide.* Amsterdam: John Benjamins Publishing Company.

Cruse, D. A. 1973. 'Thoughts on Agentivity', *Journal of Linguistics* (9)1, 11–23.

Crawford, W.J. & E. Csomay. 2016. *Doing Corpus Linguistics*. New York and London: Routledge.

Crystal, D. 2008. *A Dictionary of Linguistics and Phonetics*. Oxford: Blackwell Publishing.

Daugs, R. 2017. On the development of modals and semi–modals in American English in the 19th and 20th centuries. In: T. Hiltunen, J. McVeigh & T. Säily, eds. *Big and Rich Data in English Corpus Linguistics: Methods and Explorations (Studies in Variation, Contacts and Change in English 19)*. Helsinki: VARIENG: http://www.helsinki.fi/varieng/series/volumes/19/daugs/.

Depraetere, I. & A. Verhulst. 2008. 'Source of modality: A reassessment', *English Language and Linguistics* 12(1), 1– 25.

Glass, L. 2015. 'Strong Necessity Modals: Four Socio-pragmatic Corpus Studies', *University of Pennsylvania Working Papers in Linguistics* 21(2), 77–88.

Gruber, J. S. 1967. 'Look and See', *Language*, 43(4), 937–947.

Halliday, M. A. K. 1968. 'Notes on Transitivity and Theme in English: Part 2', *Journal of Linguistics* 3(2), 199–244.

Huddleston, R. and G.K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Jacobsson, B. 1974. 'The auxiliary need', *English Studies* (55)1, 56–63.

Johansson, S. 2013. 'Modals and semi-modals of obligation in American English: some aspects of developments from 1990 until the present day'. In: Aarts, B., Close, J., Leech, G. and S. Wallis (eds) *The Verb Phrase in English: Investigating Recent Language Change with Corpora.* Cambridge: Cambridge University Press, 372–380.

Kennedy, G. 1998. *An Introduction to Corpus Linguistics*. London and New York: Longman.

Leech, G. 2003. 'Modality on the move: The English modal auxiliaries 1961–1992'. In: *Modality in Contemporary English.* Berlin: Mouton de Gruyter, 223–240.

Leech, G., M. Hundt, C. Mair and N. Smith. 2009. *Change in Contemporary English: A Grammatical Study.* Cambridge: Cambridge University Press.

Lyons, J. 1977. *Semantics*. Cambridge: Cambridge University Press.

Matthews, P. H. 2007. *The Concise Oxford Dictionary of Linguistics.* Oxford: Oxford University Press.

Matthews, P. H. 2014. *The Concise Oxford Dictionary of Linguistics.* Oxford: Oxford University Press.

Millar, N. 2009. 'Modal verbs in TIME: Frequency changes 1923–2006*', *International Journal of Corpus Linguistics* 4(2), 191–220.

Nokkonen, S. 2015. *Changes in the Field of Obligation and Necessity in Contemporary British English: A Corpus–based Sociolinguistic Study of Semi-modal NEED TO*. Vantaa: Hansaprint.

Nuyts, J. 2014. 'Analyses of the Modal Meanings', In: *The Oxford Handbook of Modality and Mood*. Oxford: Oxford University Press, 31–49.

Palmer, F. R., 1986. *Mood and Modality.* Cambridge: Cambridge University Press.

Palmer, F. R., 1990. *Modality and the English Modals.* New York: Longman Inc.

Perkins, M. R. 1983. *Modal Expressions in English*. London: Frances Pinter.

Potts, A. & P. Baker. 2012. 'Does semantic tagging identify cultural change in British and American English?', *International Journal of Corpus Linguistics* 17(3), 295–324.

Quirk, R., S. Greenbaum, G. Leech and J. Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.

Smith, N. 2003. 'Changes in the modals and semi–modals of strong obligation and epistemic necessity in recent British English', In: Facchinetti, R., Krug, M. and F. Palmer (eds), *Modality in Contemporary English.* Berlin: Mouton de Gruyter, 241–266.

Tognini-Bonelli, E. 2001. *Corpus Linguistics at Work*. Amsterdam: John Benjamins Publishing Company.

van der Auwera, J. 1999. On the Semantic and Pragmatic polyfunctionality of modal verbs. In: Turner, K (eds), *The Semantics/Pragmatics Interface from Different Points of View.* Oxford: Elsevier, 50–64.

van der Auwera, J. & V. Plungian. 1998. 'Modality's semantic map', *Linguistic Typology* 2(1), 79–124.

Weisser, M. 2015. *Practical Corpus Linguistics : An Introduction to Corpus-Based Language Analysis.* Chichester: John Wiley & Sons, Incorporated.

Westney, P. 1995. *Modals and Periphrastics in English : An Investigation into the Semantic Correspondence between Certain English Modal Verbs and Their Periphrastic Equivalents.* Tübingen: Max Niemeyer Verlag.

Zadeh, L. A. 1972. 'A Fuzzy-set-theoretic Interpretation of Linguistic Hedges', *Journal of Cybernetics* 2(3), 4–34.

Ziegeler, D. 2020. 'Mood and Modality', In: Aarts, B., Bowie, J. and G. Popova (eds), *The Oxford Handbook of English Grammar*. Oxford: Oxford University Press, 418–438.