A Corpus-based Variationist Approach to
the Use of *It is I* and *It is Me*: A Real-time Observation
of a Syntactic Change Nearing Completion in COHA

Aimi Kuya

Fukuoka Jo Gakuin University

Abstract: This is a corpus-based variationist linguistics (CVL) study into a syntactic variation and change that is nearing completion: the shift from *it is I* (nominative) to *it is me* (accusative) in American English. Despite a number of recent statements in grammar books and dictionaries that the new variant, i.e., the accusative construction, has earned public acceptance, few attempts have been made to provide empirical support for the entire diachronic shift from nominative case to accusative case. Real-time quantitative data from the diachronic corpus COHA characterizes almost the entire process of the syntactic change, and the analysis of the data using the multivariate model demonstrates: (i) that the shift first occurred in the first-person singular *I/me*, and then in the other personal pronouns; (ii) that the shift fron *I* to *me* had been almost completed by the latter half of the 20th century; (iii) that the shift achieved a level of near-completion much earlier in contracted constructions such as *it's I/me* than in non-contracted constructions such as *it is I/me*. The present study makes a cross-disciplinary contribution to the study of language variation and change by combining advantages of variationist sociolinguistics with those of corpus linguistics. It succeeds in complementing findings from previous studies by providing an empirical, comprehensive and detailed account of the process of this well-known syntactic change.*

Key words: corpus-based variationist linguistics (CVL), sociolinguistics, language variation and change, real-time study, multiple logistic regression analysis

1. Is the Case Shift towards *It is Me* Complete?

According to Quirk et al. (1972: 210), in contemporary English, the choice of case shifts between nominative (henceforth, NOM) and accusative (henceforth, ACC) when personal pronouns function as a subject complement following the verb *to be*. This syntactic variation has long been a very controversial topic among gram-

marians. Some grammarians (e.g., Brown 1856, Onions 1904, Weiner & Hawkins 1985) have insisted on a preference for NOM over ACC at least in formal contexts, usually based on traditional grammatical "correctness" originating from Latin grammar rules (Pearsall ed. 1998: 1385). However, a number of linguists and grammarians admitted, as early as Alford (1864: §192) and Poutsma (1916: §8), among others, that in practice almost everyone used the ACC construction *it is me*. Almost 100 years ago, Sapir remarked and predicted as follows:

> The folk says *it is me*, not *it is I*, which is "correct" but just as falsely so as the *whom did you see?* that we have analyzed. *I'm the one*, *it's me; we're the ones, it's us that will win out*—such are the live parallelisms in English today. There is little doubt that *it is I* will one day be as impossible in English as *c'est je*, for *c'est moi*, is now in French. (Sapir 1921: 178–179)

The above argument clearly indicates that Sapir considered the linguistic variable in question to be involved in the process of language change. Questions arise as to (i) to what extent the case shift in the subject complement has progressed, and (ii) in what manner the shift has developed up to the present day. In this paper, first of all, a general survey will be made of grammarians' statements on the nominative-accusative alternation over the past two and a half centuries, which implies that the shift is now nearing completion. Then the above research questions will be investigated using a corpus-based variationist linguistics (CVL) approach (Szmrecsanyi 2017). Empirical evidence collected from a diachronic corpus of American written English as a case study will give a comprehensive and detailed account of the process of this linguistic change.

2. Evolution of the Use of *It is Me* as Opposed to *It is I*
2.1. Historical background of the construction *it is me*
According to Smith (1906: 77–81), the development of personal pronouns as a subject complement can be traced back to Old English (OE). As shown in Table 1, from OE to Early Middle English (Early ME) the construction used the *Ic hit eom* ('I it am') order, in which the sentence-initial element *ic* ('I') was the real subject, with the predicate *eom* ('am') agreeing with it. In Late Middle English (Late ME), *it* moved to the sentence-initial position, e.g., *It am I*, but here *it* was not perceived to be the actual subject, and the verb *am* agreed with the subsequent pronoun *I*. During "transition ME" (Smith's terminology) the verb agreed with the sentence-initial *it* as in *It is I*. This indicates that *it* became the real subject in the construction, and the case of pronouns began to be influenced by their post-verbal position. In Early Modern English (Early ModE), pronouns in the post-verbal position began to take the accusative form, and this new construction, *it is me*, emerged as a variant of *it is I*. However, there has been prescriptive resistance to the use of the ACC case in this new construction, since it was against the Latin grammatical rules for subject complements (Bauer 1998: 134).

Table 1. Historical development of personal pronouns as a subject complement (Summarized from Smith 1906: 77–81)

OE to Early ME	Late ME: English of Chaucer	Transition ME: English of Caxton	Early ModE: Tudor English; English of Shakespeare
until 1300 A.D.	1300-1400 A.D.	1400-1500 A.D.	1500-1600 A.D.
Ic hit eom	*It am I*	*It is I*	*It is me*
þū hit eart	*It art thou*	*It is thou*	*It is thee*
Hē hit is	*It is he*	*It is he*	*It is him*
Wē hit sind, etc.	*It are we*, etc.	*It is we*, etc.	*It is us*, etc.

The use of ACC case gradually became recognized despite it being considered grammatically "incorrect." Possible reasons for the transition from *it is I* (*thou/he/she/we/they*) (*it is* NOM) to *it is me* (*thee/him/her/us/them*) (*it is* ACC) include (i) sound analogy of other NOM forms: *we, ye, he, she*; (ii) syntactic analogy of sentences such as *he saw me*, *tell me*, etc.; and (iii) the influence of the French expression *c'est moi*. Sound analogy (Jespersen 1894: §193, Onions 1904: §25) explains that the accusative *me* or *thee* is a rhyme for the nominative *we, ye, he, she*. Syntactic analogy (Sweet 1891: §1085, Jespersen 1894: §184) shows that the post-verbal position of pronouns drove *it is* NOM to *it is* ACC. This hypothesis seems to be the most popular one among previous scholars including Smith (1906: 86), Curme (1931: 42), and Mustanoja (1960: 133) (see also Mencken (1919: 222) for a discussion on the strong transitiveness of the verb *to be*). The influence of the French expression *c'est moi* (Mason 1879: §459, Lounsbury 1894: §117) does not seem to gain as much support from scholars as the syntactic analogy does. For example, Smith (1906: 84) and Kisbye (1972: 103) point out a diachronic gap between French influence on English and the emergence of *it is me*.

2.2. Debate on the construction *it is me*

The variation between *it is I* and *it is me* was one of the quirks in language usage discussed by grammarians in the 18th century (Görlach 1999: 67). It seems likely that there has been a diversity in grammarians' attitudes towards the construction *it is me* since that time. Table 2 and Table 3 show statements on the use of the *it is* ACC construction in grammar books, dictionaries, and other academic publications published over the past several centuries.[1]

Two contrasting conceptions are used to categorize grammarians' statements: (i) *prescriptive* ones that show either negative evaluations on the use of ACC or positive evaluations on the use of NOM, and (ii) *descriptive* ones that try to neutrally describe or support the use of ACC in practice.

---

[1] The quotes are mainly taken from first editions. Later editions are occasionally quoted if the first edition is not easily obtainable, or if a later edition contains a new piece of information.

2.2.1. A prescriptive view (a purist view)

Table 2 lists prescriptive statements observed between the mid-18th century and the late 20th century. The famous prescriptivist Lowth (1762) simply asserts that the use of NOM is "always" correct. This gives the impression that we have no choice but to use NOM in any context. Cobbett (1819) says that pronouns "ought to" be NOM, in *it was I* (*who broke it*) because of the existence of the implicit relative clause *who broke it,* in which *I* functions as the subject. This syntactic restriction is called "relative attraction" (Jespersen 1894: §154) (this issue will be discussed further in Section 3.3). Onions (1904) also declares that the predicate pronoun "stands in the nominative case" in constructions as follows: *I am he whom you want*; *Is it we you are talking to?*, despite the presence of relative attraction. A prescriptive statement can also be found in Brown (1827), where the use of ACC is considered "not proper," and this attitude remains the same in its revised version (Brown 1856). In the 19th–20th centuries, there were many scholars who remained strictly prescriptive, stigmatizing the use of ACC as "incorrect" (Sweet 1891), "careless usage" (Curme 1931), and "only *dial.* and *vulgar*" (Murray et al. 1933). Similarly, Fowler & Fowler (1906) state that the NOM usage is "correct" and "advisable" although they mention that NOM forms sound too formal in colloquial settings. In the most recent several decades, prescriptivism in its strict sense was rarely observed but not non-existent; for example, Weiner & Hawkins (1985) declare the absolute necessity of NOM forms in formal style.

Table 2. Prescriptive statements in chronological order (Underlines and truncations ("…") are mine)

1751–1800
The Verb *to Be* has <u>always</u> a Nominative Case after it … unless it be in the Infinitive Mode (Lowth 1762: 105)
1801–1850
"Who broke the glass? It was *me*." It <u>ought to</u> be *I*; that is to say, "it was *I who broke it*." (Cobbett 1819: 98, § 185)
1851–1900
<u>Not proper</u>, because the pronoun *him* ... does not agree with the pronoun *it* … Therefore, *him* should be *he*; thus, We did not know that it was *he*. (Brown 1827: 167; 1856: 206)
... such expressions as *it is me* are still denounced as <u>incorrect</u> by the grammars, many people try to avoid them in speech as well as writing. (Sweet 1891: 341, §1085)
1901–1950
The Predicate Noun and Predicate Pronoun stand in the <u>Nominative</u> Case: It is *I*. I am *he* whom you want. (Onions 1904: 34, §25)
Many educated people feel that in saying *It is I...* instead of *It's me* … they will be talking like a book, … But in print, unless it is dialogue, <u>the correct forms are advisable</u>. (Fowler & Fowler 1906: 61)
The wide use of the accusative for the nominative, ... It is to be hoped that all who are interested in accurate expression will oppose this general drift by taking more pains to use a nominative where a nominative is in order. … It is gratifying to observe that this <u>careless usage,</u> though still common in colloquial speech, is in general less common in our best literature than it once was. (Curme 1931: 43)
Me … 6. For the *nominative*. ... now <u>only *dial.* and *vulgar*.</u> (Murray ed. 1933: 264)
1951–2000
When a personal pronoun follows *it is*, *it was*, … , etc., it <u>should always</u> have the subjective case: … Informal usage favours the objective case … This is <u>not acceptable</u> in formal usage. (Weiner & Hawkins 1985: 157)

2.2.2. A descriptive view

Table 3 shows descriptive statements on ACC constructions from the 1860s onwards. Scholars like Alford, who focused on common usage, pointed out that "every one" actually used ACC forms (Alford 1864) and even said a year later "it will be curious" to prove that the ACC variant was "good" (Alford 1865). Kellner (1892) also mentioned that ACC constructions were getting popular. Latham (1862) even speculated that the argument in favor of *I* instead of *me* was invented by grammarians.

During the first half of the 20th century, many grammar books provided positive connotations for ACC constructions. Many scholars implied that the use of ACC was "winning its way" (Smith 1906), was being recognized as "quite common" (Poutsma 1916), or had "long been natural" (Jespersen 1933). It was accepted as an established colloquial expression in a dictionary (Fowler & Fowler 1924), and Sapir (1921) implied in the 1920s that the shift would continue progressing towards completion. Mencken (1919) also stated that the use of ACC in the NOM position was observed in American English.

In the second half of the 20th century and the turn of the 21st century, the vast majority of the opinions of grammarians examined here admitted that ACC forms were "superseding" NOM forms (Potter 1975), and that the shift towards ACC had reached almost all speakers in spoken domains (Brook 1958, Vallins 1966). Greet annotates in Partridge (1963) (adapted for American publication) that ACC forms are used in "good speech" in America. Some statements imply that ACC became so "natural" (Evans & Evans 1957, Morris ed. 1969, Quirk et al. 1972), "usual" (Hornby 1974), and "common" in colloquial English (*OED Online*, updated March 2001 for the entry *me*) that the NOM counterpart sounds "pedantic"

Table 3. Descriptive statements in chronological order (Underlines and truncations ("…") are mine; an asterisk (*) shows a comment specifically on American English)

1851–1900
… is there any real custom at all in favour of *I* except so far as <u>the grammarians have made one</u>? (Latham 1862: 587, §650)
"It is me." Now this is an expression which <u>every one</u> uses. (Alford, 1864 : 142, §192)
<u>It will be curious</u> if, after all, it should be proved that our much-abused colloquial phrase is the really <u>good</u> English, and its rival "a mistaken purism." (Alford 1865: 287, NOTE F, newly added in this edition)
The expression "*it is me*" which is <u>constantly gaining ground</u>, is of comparatively recent date. (Kellner 1892: 42, §79)
1901–1950
At present it [*it is me*] is <u>winning its way</u> more rapidly than ever before, because the grammarians have begun to advocate it. (Smith 1906: 82) (square brackets mine)
This *me* is <u>quite common</u> even in the language of educated speakers. (Poutsma 1916: 715, §8)
*But against this use of the objective form in the nominative position after the verb of being there also occurs in American a use of the nominative form in the objective position, … (Mencken 1919: 222–223)
There is <u>little doubt</u> that *it is I* will one day be as impossible in English as *c'est je*, for *c'est moi*, is now in French. (Sapir 1921: 179)
me … also used … subjectively (colloq.; *it's me*) (Fowler & Fowler 1924: 488)
It is and has <u>long been natural</u> to use the objective in the predicative (Jespersen 1933: 135, §14.3.4)

1951–2000

*me. In <u>natural</u>, well-bred English, *me* and not *I* is the form of the pronoun used after any verb, even the verb *to be*. When mayor Cermak of Chicago was shot by a bullet intended for Franklin Roosevelt, he said: "I'm glad it was *me* instead of you." A local newspaper thought they could improve the dying man's words and quoted him as saying, "I'm glad it was I." (Evans & Evans 1957: 294)

One of the most frequently discussed problems is whether to say *It is I* or *It is me*. The latter expression gained ground so quickly that is now the <u>usual</u> idiom, especially in colloquial speech. (Brook 1958: 152)

... in the present stage of the battle, most people would think 'it is I' <u>pedantic</u> in talk, and 'it is me' improper in writing. (Gowers 1962: 198)

*[In America, *it's me* is acceptable colloquial English; that is, it is used in <u>good speech</u>. There is no occasion to write it. *Us, him*, *her*, *them* are less common after *to be*, and their acceptableness is disputed.] (Partridge 1963: 160, [ ]: annotated by Greet)

Written English as a rule prefers the nominative forms (*she saw that it was he*), though *it is I* is often felt to be <u>pedantic</u>. (Zandvoort 1965: 129, §360)

... we hear in the speech of <u>all classes</u> of society such expressions as "it was me", … , perhaps <u>more frequently than the prescribed form</u>. (Vallins 1966: 133)

***me** … *I*, rather than *me*, is the grammatically prescribed first person pronoun for use after the verb *be*: *It is I*. In formal writing, *it is I* is the construction specidied by 78 per cent of the Usage Panel. The variant *it is me* (or *it's me*) is felt by many persons to be much more <u>natural</u> in speech, and this form is termed acceptable in speech on all levels by 60 per cent of the Panel. (Morris ed. 1969: 810)

Although the prescriptive grammar tradition stipulates the subjective case form, the objective case form is normally felt to be the <u>natural</u> one, particularly in informal style. (Quirk et al. 1972: 210, §4.112)

me … *It's me* (now <u>usu</u> for 'it is I'). (Hornby et al. 1974: 534)

... traditional grammar tells us that the subjective form is the correct form to use. But in practice, the subjective form sounds rather <u>stilted</u>, and is <u>avoided</u>. (Leech & Svartvik 1975: 261, §687)

Clearly the time has come for us to recognize that objective forms are gradually <u>superseding</u> subjective ones (Potter 1975: 149)

In informal English (for instance, ordinary conversation) we use object-forms (*me, him*, etc) after *be*. … In more formal English, subject-forms are possible after *be*. … However, these are very formal and <u>unusual</u> ... (Swan 1980: §135.1)

This usage is now changing even in formal English, and in informal English, the object form of the pronoun is <u>definitely preferred</u>, (Celce-Murcia & Larsen-Freeman 1983: 123)

In informal English, that is, the objective pronoun is the <u>unmarked</u> case from, used in the absence of positive reasons for using the subjective form. (Quirk et al. 1985: 337–338, §6.5)

***me** … 2. (used instead of the pronorn *I* in the predicate after verb *to be*): *It's me.* … Today, such constructions—*It's me. That's him. It must be them*—are almost <u>universal</u> in informal speech. In formal speech and in edited writing, however, the subjective forms are used: *It must be they.* ... (Costello ed. 1991: 839)

The general <u>pedantic</u> connotation is now being admitted by some grammarians. (Wales 1996: 95)

... despite the objections of prescriptive grammarians (whose arguments are based on Latin rather than English), it is <u>standard accepted</u> English to use any of the following: *Who is it? It's **me***! (Pearsall ed. 1998: 1385)

... accusative forms are <u>predominant in all registers</u>…In conversation,…, they are <u>nearly universal</u>. (Biber et al. 1999: 335–336, §4.10.6.1)

2001–

Me ... 5. For the subjective pronoun *I*. Several of these uses ... , while <u>common</u> in colloquial English, have been regarded as nonstandard by many grammarians since the 18th cent. (*OED Online*, updated March 2001 for the entry)

***me** … used after the verb "to be"… *hi, it's me*… (Jewell & Abate eds. 2001: 1058)

In most, the nominative is <u>restricted to formal (or very formal) style</u>, with the accusative appearing elsewhere. (Huddleston & Pullum 2002: 459, §16.2.1)

(Gowers 1962, Zandvoort 1965, Wales 1996), "stilted" (Leech & Svartvik 1975), or "unusual" (Swan 1980). Huddleston & Pullum (2002) see that the use of NOM forms is restricted only to "formal (or very formal)" style. In other words, the use of ACC forms in informal style is "definitely preferred" (Celce-Murcia & Larsen-Freeman 1983), "unmarked" (Quirk et al. 1985), "almost universal" (Costello ed. 1991), and "nearly universal" (Biber et al. 1999). Pearsall ed. (1998) and Jewell & Abate eds. (2001) do not make any mention of stylistic restrictions on ACC usage.

2.2.3. Conflict between prescriptivism and descriptivism

The above conflict between prescriptive and descriptive statements demonstrates "variation" and "fuzziness" (Aitchison 1981: 49–53) in attitudes towards case alternation. Note that the tables include a range of publication types varying from those aimed at linguists (e.g., Curme 1931, Jespersen 1933) to those written especially for English learners or instructors (e.g., Swan 1980, Celce-Murcia & Larsen-Freeman 1983). However, such a distinction is not always apparent and may even become fuzzier as a publication gets older. In 18th-century publications like Lowth (1762), for example, it was often the case that grammar was identified with school/purist grammar, since it was a while before linguistics came to be considered a "descriptive" science as Saussure (1857–1913) emphasizes (Lyons 1968: 42).

Some might point out that the majority of the publications investigated in this section are those of British-based grammarians and cannot be generalized to American English. This might be in a sense true; however, one should note that statements written by British grammarians do not necessarily reflect their attitudes exclusively towards the use of their own variety of English. In the same sense, it cannot be assumed that those written by American-based grammarians are exclusively describing American English, unless the author clearly states which variety of English s/he is describing (for the comments specifically on American English, see those marked with an asterisk in Table 3: Mencken (1919), the annotation by Greet in Partridge (1963), Morris ed. (1969), Costello ed. (1991), etc.).

The present paper therefore considers statements from dictionaries and grammar books to be interpreted as general descriptions of English overall. That having been said, what can be drawn from the overview of grammarians' statements on the use of ACC case is that the general tendency has changed from rejection to acceptance over the past two and a half centuries. The conflict between prescriptivism and descriptivism seems to be most remarkable from the mid-19th century to the mid-20th century, where grammatical explanations of case alternation became more varied. During this period, strong prescriptivism was still preserved in some grammar books, e.g., Curme (1931: 43), although the author admitted "the wide use of accusatives" in practice. Prescriptive resistance to ACC usage can also be found in a local newspaper: Mayor Cermak of Chicago was quoted as saying *I'm glad it was I* instead of his actual saying *I'm glad it was me* when he was shot in the attempted assassination of Franklin Roosevelt in 1933 (reported in Evans & Evans 1957: 294). Towards the turn of the 21st century, it becomes harder to find

prescriptive statements. It is worth mentioning that OED changed its statement on the use of *me* as a subject complement from "only *dial.* and *vulgar*" in its 1933 edition (Murray et al. 1933) as shown in Table 2 to "common in colloquial speech" in its online edition (updated March 2001) as shown in Table 3.

3. Analytical Framework

3.1. Corpus-based variationist linguistics (CVL)

The current study adopts a corpus-based variationist linguistics (henceforth, CVL) approach. According to Szmrecsanyi (2017), CVL is differentiated from corpus linguistics and variationist sociolinguistics (also known as the Language Variation and Change or LVC), although they are all within the scope of empirical linguistics as illustrated in Figure 1.

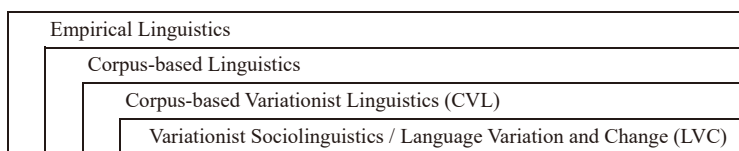| Empirical Linguistics			
	Corpus-based Linguistics		
		Corpus-based Variationist Linguistics (CVL)	
			Variationist Sociolinguistics / Language Variation and Change (LVC)

Figure 1. How CVL is related to other branches of empirical linguistics (Adapted from Szmrecsanyi 2017: 687, Figure 1)

In one sense, CVL is a subset of corpus linguistics; in other words, "not all corpus-based research is concerned with linguistic variation in the variationist sense" (Szmrecsanyi 2017: 687). Szmrecsanyi states that CVL is concerned with (i) the notion of linguistic variables, i.e., several different ways (the option) of saying "the same thing" (Labov 1972: 271), (ii) the Principle of Accountability (Labov 1972: 72), and (iii) rigorous quantitative methodologies and statistical modeling. In another sense, CVL is defined as a superset of variationist sociolinguistics, i.e., variationist sociolinguistics is "a particular way of doing CVL" provided that data collected with traditional sociolinguistic interviews fall into the category of corpora (Szmrecsanyi 2017).

The two fields of study, CVL and variationist sociolinguistics, are different in the following six aspects (Szmrecsanyi 2017: 688–689). CVL tends to put more emphasis on (i) macro-sociological factors, e.g., "economization" (the tendency towards informational compression) and "colloquialization" (use of informal language), (ii) non-phonetic (in particular grammatical) variation, (iii) real-time (diachronic) approach,[2] (iv) existing, publicly available corpora with (semi-) automatic retrieval and annotation procedures, which result in a big-data project,

---

[2] The term *real-time* approach is used to contrast with the term *apparent-time* approach in the context of variationist sociolinguistics (also known as *diachronic* approach as opposed to *synchronic* approach). The former refers to methods that directly observe linguistic change over time; whereas the latter refers to methods that indirectly observe linguistic change over time by utilizing "the distribution of linguistic forms across age groups" (Labov 1994: 28). The advantages of adopting a real-time approach for this study are discussed in Section 5.

(v) innovative analytic techniques (e.g., multinominal logistic regression analysis[3]), and (vi) the view of variation analysis as an exercise in usage-based linguistics (for usage-/corpus-based approaches to grammar, see, for example, Biber et al. 1999, Bybee 2006, and Miller 2011). In contrast, traditional variationist sociolinguistics usually tends to be more interested in (i) demographic factors (e.g., speaker age, gender, etc.), (ii) phonetic variation, (iii) apparent-time (synchronic) approach, (iv) fieldwork with manual annotation, (v) plain binary logistic regression, and (vi) a more controversial view of variation analysis as an exercise in usage-based linguistics. Szmrecsanyi (2017: 686) sees that work in CVL has been inspired by work in variationist sociolinguistics and that methodologies and research questions of CVL will "fuel" work not only in variationist sociolinguistics but also in corpus linguistics.

In the above-mentioned sense, the present study is defined as being under the field of CVL in that it investigates syntactic variation using a variationist approach consisting of the statistical modeling of language change based on a real-time (diachronic) observation of data collected from an existing, large corpus using an automatic retrieval technique. The focus is, as illustrated in what follows, on the effect of more macroscopic factors (publication year and register) rather than demographic factors, and grammar is seen as what is based upon usage and experience.

3.2. Research questions

Alford (1865: 285–286, NOTE F) emphasizes an empirical point of view on the phenomenon by quoting Ellis' statement in his letter to *The Reader*, of May 7, 1864: "I consider that the phrase *it is I* is a modernism, or rather a grammaticism—that is, it was never in popular use, but was introduced solely on some grammatical hypothesis as to having the same case before and after the verb *is*." Ellis raised the question "is the prescriptive grammar a mere hypothesis?" As we have seen, researchers and grammarians with a descriptive view on the phenomenon have shown that it is not. For example, Jespersen (1894: §184) provides plenty of examples of NOM constructions used by writers like Chaucer (1343–1400) and Shakespeare (1564–1616), e.g., 'it was *he*,' or ''tis *hee*' (Shak., *Macb.*). Smith (1906: 82) further states that Shakespeare was "overwhelmingly in favor of" NOM forms. Dekeyser (1975: 224) and Nakayama (2014) show that the *it is I* construction was still predominant in 19th-century British English. The use of ACC was considered widespread by the 19–20th centuries (e.g., Alford 1864: 142, Jespersen 1933: 135, among others). American English was no exception (Mencken 1919: 222–223, Evans & Evans 1957: 294, Partridge 1963: 160, Morris ed. 1969: 810). Around the turn of the 21st century it is said that the shift towards ACC constructions was nearing completion (e.g., Biber et al. 1999: 336, Jewell & Abate eds. 2001: 1058, among others).

---

[3]   See, for example, Han et al. (2013) for the treatment of five Shanghainese topic markers as a dependent variable. Other terminologies for *multinominal* logistic regression include *polychotomous*/*polytomous* logistic regression (Hosmer et al. 2013: 269).

These empirical findings, as well as the statements in earlier publications discussed in Section 2.2, suggest the following hypothesis: the use of *it is* ACC surpassed *it is* NOM at some point during the 19th–20th centuries and the shift has now almost been completed. In previous studies, however, not much emphasis has been put on diachronic research that observes the entire process of this syntactic change. In order to fully verify the hypothesis above, a real-time observation of data across at least two centuries will be required. Research questions for the current study of American English arise then, as to (i) how far the shift has progressed over the past two centuries, and (ii) in what manner the new variant diffused during that period. An investigation into "the language of the folk," to use Sapir's expression (1921: 178), would contribute to the description of usage-based grammar.

3.3. Scope of analysis

It has been pointed out that the effect of "relative attraction" (Jespersen 1894: §154) restricts the choice of case. When a personal pronoun is followed by a relative clause in which it functions as the subject as in (i) in Table 4, NOM is preferred. Based on a corpus survey, Biber et al. (1999: §4.10.6.1) report that NOM forms usually appear before a *who*-clause. Erdmann (1978: 76)[4] also reports that of the 119 examples preceded by a *who*-clause, 99 are NOM forms in British novels published between 1930 and 1980. In contrast, when a personal pronoun is seen to be the object in the relative clause following it as in (ii), ACC is preferred (see also Curme 1931: 42, Visser 1970: §265). When a personal pronoun is not followed by a relative clause as in (iii), "it is and has long been natural to use the objective in the predicate" (Jespersen 1933: §14.3.4).

Table 4. Syntactic structure and stylistic "flavour" (Adapted from Huddleston & Pullum 2002: 459)

	NOMINATIVE	ACCUSATIVE
i	a. *It is I who love you*.	b. *It's me who love you.* [informal]
ii	a. *It is I she loves*. [very formal]	b. *It's me she loves.* [neutral]
iii	a. *It is she!* [very formal/pedantic]	b. *It's her!*

Huddleston & Pullum (2002: 459) accept the use of both NOM and ACC cases in the constructions (i)–(iii) in contemporary English, but still, they attract our attention to possible stylistic differences between contrasting sentences (a and b). In the first construction (i), the use of NOM (a) is unmarked in style whereas the use of ACC (b) "certainly has an informal flavour." In the second construction (ii), ACC (b) sounds "relatively neutral in style" whereas its NOM counterpart (a) may strike some people as "very formal." In the third construction (iii), the use of NOM (a) is considered "very formal" or "pedantic" in response to the question

---

[4] A source suggested by a reviewer.

Who's there? and that of ACC (b) is unmarked in style. In other words, prescriptive resistance to ACC usage no longer seems to bother contemporary English speakers.

It is apparent that the perceived stylistic distinctions, i.e., stylistic "flavour" in the terminology of Huddleston & Pullum (2002), are due to syntactic restrictions imposed on the case selection. The existence of a relative clause in (i) and (ii) causes relative attraction and affects the choice of case, and the sentences are often judged negatively in terms of style ("informal," "very formal," "pedantic") when the grammatically disfavored variants are chosen. An analytical problem for the current study is that syntactic restrictions caused by relative attraction are overt in the constructions (i) and (ii), whereas they are not overt in the structures (iii). In other words, a construction that does not include a relative clause demonstrates stronger fuzziness to be included in variation analysis. Therefore, the scope of analysis of the present study will be limited to constructions without relative attraction as in (iii) in Table 4, and constructions as in (i) and (ii) are excluded. The case shift is probably progressing at a slower rate in contexts where relative attraction occurs than in those where it does not occur, but a comparison of the two will be left for future investigation.

4. Empirical Evidence from COHA: A Case Study
4.1. Collecting real-time data
CVL methodologies were employed to collect empirical evidence for the above-mentioned hypothesis that ACC usage surpassed NOM usage during the 19th–20th centuries. The 400-million-word diachronic corpus, the Corpus of Historical American English (COHA) in the BYU corpora created by Mark Davies at Brigham Young University, was utilized in this section to conduct a case study of "the language of the folk" (Sapir 1921: 178). COHA consists of written language from the 1810s to 2000s across four genres: FICTION (FIC, 51.1%), MAGAZINES (MAG, 23.9%), NEWSPAPERS (NEWS, 9.9%), and NON-FICTION BOOKS (NF, 15.1%) (see Table 6 for the total word count for each genre and decade). The corpus is generally "balanced by genre across the decades" (Davies 2012: 123), but attention should be paid to the fact that NEWS includes data only from the 1860s to 2000s (see Davies (2012: Table 1) for details of the composition of COHA).

Findings from the investigation of this corpus may not be immediately generalized to all English varieties; nevertheless, this corpus is very beneficial to the present study in that it is an existing, publicly available diachronic corpus that covers a relatively long period: two hundred years from 1810 to 2009. This generally corresponds with the period that the above-mentioned hypothesis is mainly concerned with (Section 3.2). Real-time observation of this corpus will be useful for verifying the hypothesis regarding the syntactic change that has long been well known but has not yet been fully endorsed by empirical evidence.

Data was retrieved from COHA using the interface of the BYU corpora (KWIC display). For the reason mentioned in Section 3.3, the construction

examined here is not followed by any relative clause, i.e., [*it* + *is/was/were* + PERSONAL PRONOUN] at the end of a sentence, and is therefore least likely to be affected by relative attraction. The end of a sentence was identified by the current author as a "primary terminal" (Huddleston & Pullum 2002: 1731), otherwise known as a punctuation mark "./!/?". Of the three primary terminals, tokens of "." and "?" were collected by employing the part-of-speech tag [y*] ('symbols') used in the BYU interface.[5] Other irrelevant symbols under this category (such as ( ) , ; :) were excluded afterward. A search for "!" had to be made separately since this symbol was not tagged as [y*] in COHA.

Table 5 shows the constructions considered and the corresponding search queries: the basic *it is I/me* constructions, its contracted constructions (*it's I/me*), negative constructions (*it is* (*it's*) *not I/me*), and constructions with a contracted negative (*it isn't I/me*). All other personal pronouns that maintain the NOM-ACC case distinction in Late ModE, including *he/him, she/her, we/us, they/them* are examined; however, *thou/thee* are not within the scope of this paper because they were replaced by *you* towards the end of the 17th century (Barber 1976: 208).

Table 5. Examples of constructions considered

Construction	Search queries
it is I/me	*it* [vb*] *I, it* [vb*] *me*, etc.
it's I/me	(*is, was, were,* and *'s* are subcategories of [vb*])
it is (*it's*) *not I/me*	*it* [vb*] *not I, it* [vb*] *not me*, etc.
it isn't I/me	*it* [vb*] *n't I, it* [vb*] *n't me*, etc.

The actual samples of retrieved tokens include the following: *"It's me. Open the door," said a familiar voice* (1949 FIC, *Brave Bulls*); *But of course she knew it was him.* (2001 FIC, *Riptide*); *Yes; a thrice-sodden fool – if it were I!* (1839 FIC, *Dying Keep Him*). All the tokens collected using an automatic retrieval technique with the search queries in Table 5 were checked by the current author and a couple of irrelevant tokens such as *so the only person* [*able to open it*] *was me.* (square brackets mine) (1996 MAG, *Smithsonian*) were excluded manually. It should be noted that the search queries in Table 5 do not contain constructions with an element such as an auxiliary verb between *it* and *to be* (e.g., *it will be me*) and constructions with the contracted negative *to be*, i.e., *ain't* (e.g., *it ain't me*).[6] Although the use of

---

[5] As a reviewer points out, relying on parts of speech tagged by an automatic word-tagging system in the interface could be a questionable technique. However, this does not seem applicable to the present research. The number of tokens of "./?" retrieved from the entire COHA as a result of using [y*] was identical to the number retrieved by directly searching for the individual symbols "./?".

[6] Suggested by a reviewer. Attention also needs to be paid to the possibility of irrelevant tokens such as *ME* (Middle English) or *US* (United States) mixed in the search results as a result that a search for *me/us* is case insensitive (another suggestion made by the same reviewer); however, such examples were not detected in the current data.

automatic data retrieval procedures with the search queries above is not completely unproblematic, for the purpose of the present research, the results suggest that it succeeded in collecting at least typical patterns of the *it is I/me* construction in which the NOM-ACC alternation occurs.

4.2. General trends over time

As a result of the above-mentioned automatic retrieval of data followed by manual removal of irrelevant tokens, 1,489 target tokens were left for quantitative analysis. Table 6 shows the raw frequencies of NOM and ACC according to genre (GENR) and year of publication (YEAR) (the total word count for each decade and genre is also shown in million words [mw]). It shows that FIC contributed 362 NOMs and 1,039 ACCs (= 1,401 tokens in total), which equals 94.1% of the total tokens collected, followed by MAG (14 NOMs + 41 ACCs = 55 tokens, 3.7% of the collected tokens), NEWS (2 NOMs + 17 ACCs = 19 tokens, 1.3% of the collected tokens), and NF (8 NOMs + 6 ACCs = 14 tokens, 0.9% of the collected tokens). It is surprising that almost all tokens (94.1%) come from FIC, considering that FIC contributes only 51.1% of the entire COHA word count.

Figure 2 is a graphical representation of the proportion of ACC use to NOM use based on the information listed in the row named ALL GENR in Table 6. The figure captures almost the entire process of the shift from NOM to ACC. It clearly demonstrates a constant increase in the proportion of ACC to NOM from below 10% in the 1820s to above 90% in the 1970s onward, with the use of ACC forms surpassing that of its NOM counterparts in the 1900s. The high percentage of ACC shown in the 1810s is caused by the tiny data sample of only two tokens and can therefore be considered skewed and almost definitely unrepresentative of the language used during that period.

Table 6. The number of collected NOM and ACC tokens by GENR and YEAR (COHA)

| YEAR	1810	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	Total
	1mw	7mw	14mw	16mw	16mw	17mw	19mw	21mw	21mw	23mw	23mw	26mw	24mw	24mw	24mw	24mw	24mw	25mw	28mw	29mw	406mw
FIC (208mw)																					
NOM	1	27	26	31	26	34	24	21	24	32	14	12	17	22	11	6	10	10	8	6	362
ACC	1	2	4	7	7	22	20	13	19	36	29	38	51	58	75	60	112	111	205	169	1039
MAG (97mw)																					
NOM			1	1		1		2	2	3		1				3					14
ACC						1				1		2	2	3	1	9	6	4	5	7	41
NEWS (40mw)																					
MON							1	1													2
ACC															1	1	1		1	3	10	17
NF (61mw)																					
NOM				1	1			1	1	2						1				1	8
ACC			1	1						1						1			1		1	6
ALL GENR (406mw)																					
NOM	1	27	26	32	28	35	25	21	28	36	19	12	18	22	11	10	10	10	8	7	386
ACC	1	2	5	8	7	22	21	13	19	37	30	40	53	62	78	70	118	117	213	187	1103
Total	2	29	31	40	35	57	46	34	47	73	49	52	71	84	89	80	128	127	221	194	1489
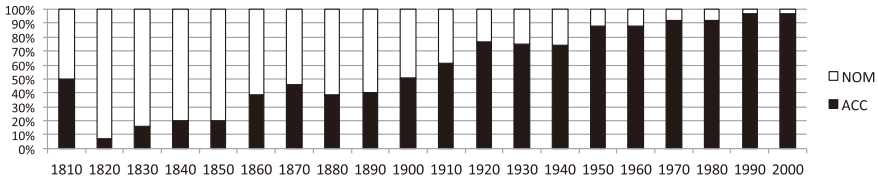
Figure 2. Diachronic growth in the use of ACC by YEAR (ALL GENR)

4.2.1. Genre of text (register)

Figure 3 visualizes a diachronic change in the proportion of ACC to that of NOM according to genre (GENR) based on Table 6. Since NEWS and NF have too low a frequency to be divided into 10-year intervals, 50-year intervals were utilized to make general tendencies over time more visible (Period I: the 1810s–1850s, Period II: the 1860s–1900s, Period III: the 1910s–1950s, and Period IV: the 1960s–2000s). The same applies to the analyses in Sections 4.2.2–4.2.4. Note that the focus of this section is just to provide an overview of diachronic distributions. Discussions of statistical significance of the effect of a given independent variable will be provided later in Section 4.3, where multivariate analysis measures the effect of the independent variable when the other variables remain constant.

The figure shows a constant shift towards ACC in FIC and MAG, reaching 50% in Period III and then 90% in Period IV. NF also shows an increase in the use of ACC in Periods II–IV with an irregular distribution in Period I, but this is not necessarily representative of actual usage since NF has a very small number of tokens compared to FIC and MAG. For the same reason, the data distribution in NEWS should be interpreted with care.
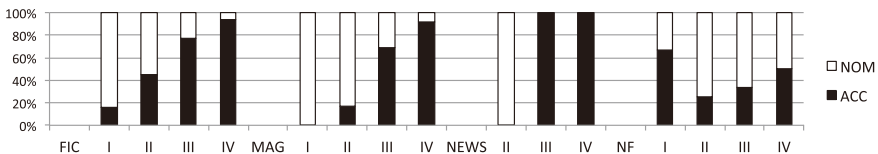

Figure 3. Diachronic growth in the use of ACC by GENR

4.2.2. Type of pronouns (*I/me, he/him, she/her, we/us, they/them*)

Sweet (1891: §1085) points out: "in standard spoken English the absolute use of the objective forms is most marked in the case of *me*, which is put on a level with the old nominatives *he*, etc.: *it is me, it is he, it is she*." Onions (1904: §25) states that *it's me* is currently used so frequently that even educated speakers use it, but the use of other pronouns in ACC case generally sounds "vulgar or dialectal." Greet also says in Partridge (1963: 160) that personal pronouns in ACC case other than *me* are less common after *to be* in American English (see Table 3).

These apparently impressionistic statements on the popularity of *it is me* as

opposed to *it is him, it is her*, etc., however, have not been verified quantitatively. Biber et al. (1999: §4.10.5) state that in conversation *me* shows the highest frequency per million words among personal pronouns (with *us* having the lowest frequency). However, this does not necessarily mean that the proportion of ACC to NOM after the verb *to be* is highest for the first-person singular. The effect of type of pronoun on the syntactic change therefore needs to be verified from a variationist viewpoint based on the relative frequency of the new ACC variant as opposed to the competing NOM variant, not based solely on the absolute frequency of the new ACC structure.

Table 7 indicates the distribution of NOM and ACC according to type of pronouns (PRON). Among the ACC pronouns, *me* is highest in raw frequency: 667 out of 1,103. This accounts for 60% of all ACC tokens. Figure 4 shows that the shift towards ACC progressed throughout all periods for all pronoun types, with the proportion of *me* in the variationist sense being highest overall. In other words, the shift towards ACC case has been pioneered by the first-person singular. Both the relative commonness of *I/me* and the fact that *I/me* changed first explains why grammarians noticed the more frequent use of *me* before the use of other pronouns in ACC case.

Table 7. Distribution of NOM and ACC by type of pronouns (PRON)

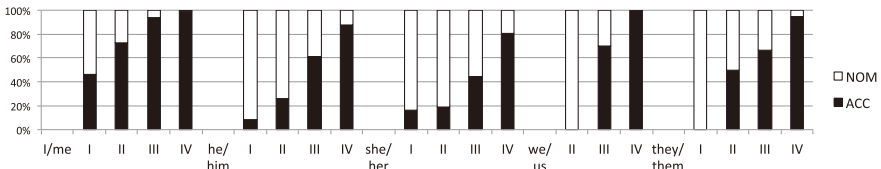| PRON	*I/me*					*he/him*					*she/her*					*we/us*				*they/them*					
Period	I	II	III	IV	Sub-total	I	II	III	IV	Sub-total	I	II	III	IV	Sub-total	II	III	IV	Sub-total	I	II	III	IV	Sub-total	Total
NOM	13	28	11	4	56	75	74	37	20	206	25	39	26	19	109	1	3		4	1	3	5	2	11	386
ACC	11		74	167	415	667	7	26	58	148	239	5	9	21	78	113		7	27	34		3	10	37	50	1103
Total	24	102	178	419	723	82	100	95	168	445	30	48	47	97	222	1	10	27	38	1	6	15	39	61	1489

Figure 4. Diachronic growth in the use of ACC by PRON

4.2.3. Negation (negative vs. non-negative)

Wales (1996: 95) argues that NOM in negative constructions (e.g., *it wasn't I/he/they*) sounds even more pedantic than plain NOM constructions (e.g., *it was I/he/they*). This observation implies that ACC is favored more in negative constructions than in non-negative ones.

Table 8 shows the distribution of NOM and ACC in the presence/absence of negation (NEG/non-NEG). Figure 5 demonstrates that both negative (NEG) and non-negative (non-NEG) constructions show a diachronic growth in the use of ACC. The proportion of ACC appears to be slightly higher in NEG construc-

tions than in non-NEG constructions, but it is not immediately evident. This will be resolved later with multivariate analysis.

Table 8. Distribution of NOM and ACC in non-NEG/NEG

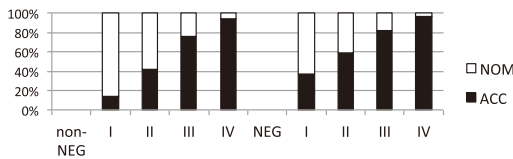|	non-NEG					NEG					
Period	I	II	III	IV	Sub-total	I	II	III	IV	Sub-total	Total
NOM	102	133	73	40	348	12	12	9	5	38	386
ACC	16	95	223	589	923	7	17	40	116	180	1103
Total	118	228	296	629	1271	19	29	49	121	218	1489

Figure 5. Diachronic growth in the use of ACC in non-NEG/NEG

4.2.4. Contraction (contracted vs. non-contracted)

One of Minami's (2009: 189–190) informants stated that they would have chosen *I* if the personal pronoun follows *it is* instead of the contracted construction *it's*. Nakayama (2014: 5) also showed the tendency for *me* to appear more frequently in contracted constructions, but the statistical significance of this trend was not tested.

Table 9 demonstrates the distribution of NOM and ACC in the presence/absence of contraction (CONT/non-CONT). Figure 6 clearly indicates that a diachronic growth of ACC usage is observable for both constructions, and that the occurrence of ACC has been frequent where contractions occur (CONT) for the entire period examined here. Table 10 and Figure 7 confirm that all types of contraction (i.e., *it's*, *it isn't*, *it's not*) have maintained relatively high frequencies of ACC throughout the entire period (the irregularity for the predicate *it is not* in Period III is probably due to an insufficient number of tokens). It is hypothesized that the shift first occurred in contracted constructions and diffused later into non-contracted constructions.

Table 9. Distribution of NOM and ACC in non-CONT/CONT

	non-CONT					CONT					
Period	I	II	III	IV	Sub-total	I	II	III	IV	Sub-total	Total
NOM	109	136	70	36	351	5	9	12	9	35	386
ACC	14	38	85	187	324	9	74	178	518	779	1103
Total	123	174	155	223	675	14	83	190	527	814	1489

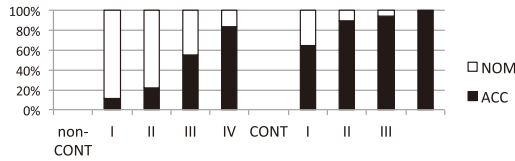Figure 6. Diachronic growth in the use of ACC in non-CONT/CONT

Table 10. Distribution of NOM and ACC by type of predicate

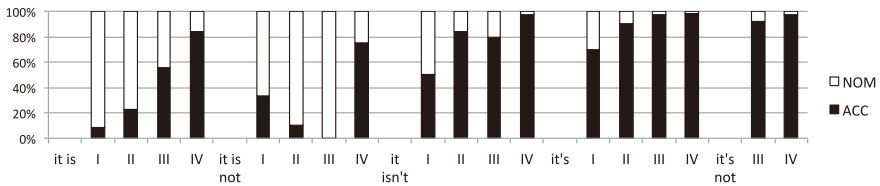| Period	*it is*					*it is not*					*it isn't*					*it's*					*it's not*			Total
	I	II	III	IV	Sub-total	I	II	III	IV	Sub-total	I	II	III	IV	Sub-total	I	II	III	IV	Sub-total	III	IV	Sub-total	
NOM	99	127	69	34	329	10	9	1	2	22	2	3	7	2	14	3	6	4	6	19	1	1	2	386
ACC	9	37	85	181	312	5	1		6	12	2	16	28	75	121	7	58	138	408	611	12	35	47	1103
Total	108	164	154	215	641	15	10	1	8	34	4	19	35	77	135	10	64	142	414	630	13	36	49	1489

Figure 7. Diachronic growth in the use of ACC by type of predicate

4.3. Statistical modeling of the change towards the *it is* ACC construction

In this section the multiple logistic regression analysis was employed in order to evaluate the statistical significance of given individual independent variables (with the other variables being held stable) and the extent to which they affect the occurrence of the ACC variants. The dependent variable is categorical: occurrence of an ACC variant (*me/him/her/us/them*) (=1) or non-occurrence of it, i.e., occurrence of a NOM variant (*I/he/she/we/they*) (=0). The probability (*p*) of ACC variants occurring as against their NOM counterparts was calculated with the following equation that includes multiple independent variables (x_n):

$$\log[p/(1-p)] = a_1 x_1 + a_2 x_2 + \ldots + a_n x_n + b \text{ [Equation 1]}$$

Based on the discussions in Section 4.2, both language-internal and language-external variables were included in the analysis (Table 11). Publication year (YEAR) (x_1) is the most important variable in identifying the upward trend in the use of ACC over time. Genre (GENR) (x_2) and the three language-internal factors: type of personal pronoun (PRON) (x_3), negation (NEG) (x_4), and contraction (CONT) (x_5) were included in the analysis to examine their statistical significance. YEAR is a numerical variable (10-year intervals are used for this multivariate analysis) while the rest of the variables are categorical. The category coded with the smallest value is the reference for each variable, i.e., the reference category is FIC

for the variable GENR; *I/me* for the variable PRON; *is/'s* for the variable NEG; *is/is not* for the variable CONT.

Table 11. Independent variables (x_n) tested for statistical significance

x_1	YEAR	(numerical)	1810s =1810, 1820s =1820, 1830s = 1830, … 2000s = 2000
x_2	GENR	(categorical)	FIC = 1, MAG = 2, NEWS = 3, NF = 4
x_3	PRON	(categorical)	*I/me* = 1, *he/him* = 2, *she/her* = 3, *we/us* = 4, *they/them* = 5
x_4	NEG	(categorical)	*is/'s* = 0, *is not/'s not/isn't* = 1
x_5	CONT	(categorical)	*is/is not* = 0, *'s/'s not/isn't* = 1

The results of the logistic regression analysis conducted by SPSS (stepwise, at the 5% level of significance) are shown in Table 12. Variables that did not reach statistical significance were excluded from the table. A positive coefficient (B), which corresponds with a_n in Equation 1, indicates that the variable in question has a positive impact on the preference for ACC over NOM, and a negative coefficient means the reverse. Constant in Table 12 corresponds with b in Equation 1.

Table 12. Results of the logistic regression analysis

Variables in the model	B	S.E.	Wald	df	Sig.	Exp(B)
YEAR (x_1)	0.025227	0.002	184.865	1	$p < 0.001$	1.026
PRON (x_3)			96.773	4	$p < 0.001$	
he/him	-1.886789	0.219	74.500	1	$p < 0.001$	0.152
she/her	-2.199510	0.260	71.735	1	$p < 0.001$	0.111
we/us	-1.270000	0.643	3.903	1	0.048	0.281
they/them	-0.980329	0.443	4.896	1	0.027	0.375
CONT (x_5)	2.435510	0.215	128.601	1	$p < 0.001$	11.422
Constant	-47.196372	3.546	177.163	1	$p < 0.001$	0.000

Nagelkerke R^2 = 0.653

The table above demonstrates that YEAR, PRON, and CONT were shown to be significant, whereas GENR and NEG did not reach statistical significance and were excluded from the model. First of all, YEAR has a positive coefficient, confirming a diachronic increase in the proportion of ACC to that of NOM. Second, negative coefficients for *he/him*, *she/her*, *we/us*, and *they/them* in PRON show that these pronouns favor ACC less than the reference category *I/me* does. This confirms that the change towards ACC was pioneered by the first-person singular *me*, which provides statistical evidence for the observations of Sweet (1891), Smith (1906), and Partridge (1963) in the variationist sense. Third, the positive coefficient for CONT indicates that contractions encourage the use of ACC, providing statistical evidence to support the observations of Minami (2009) and Nakayama (2014). The odds ratio, represented by Exp(B), indicates

that that the odds $p/(1-p)$ for ACC variants occurring in a construction with a contraction (CONT) is 11.422 times as high as the odds for the same variants occurring in a construction without a contraction (non-CONT). Finally, non-significance of GENR is probably due to the imbalance in genre distribution in the current data (i.e., the predominance of data from FIC), and it does not necessarily mean that genre is unimportant in predicting the occurrence of ACC.

Figure 8 compares the rate of the shift towards the use of *me* instead of *I* after *it is/it is not* (solid line) with that after their contracted constructions: *it's/it's not/it isn't* (dotted line) predicted by the logistic regression model (Table 12). The equation for the probability of *me* occurring in non-CONT constructions in the 1900s, for instance, is $\log[p/(1-p)] = 0.025227(1900) + 2.435510(0) - 47.196372$; while the equation for the probability of *me* occurring in CONT constructions in the same decade is $\log[p/(1-p)] = 0.025227(1900) + 2.435510(1) - 47.196372$. The model in Figure 8 demonstrates that the case shift from NOM to ACC had already reached approximately 90%, a level of near-completion, for both sentence constructions (non-CONT and CONT) by the latter half of the 20th century. The shift towards ACC in CONT constructions spread into the language nearly 100 years earlier than that in non-CONT constructions did; for example, the former reached 90% as early as the 1860s–1870s but the latter did the same in the 1960s. Note that this statistical model allows one to predict the diffusion of *me* back to the 16th century, when ACC variants were reported to have begun to spread slowly (see Smith 1906: 81–82). The prediction of the model shows that in American written English the percentage of ACC usage had remained less than 10% until the middle of the 17th century but accelerated in the 18th century for CONT and in the 19th century for non-CONT.
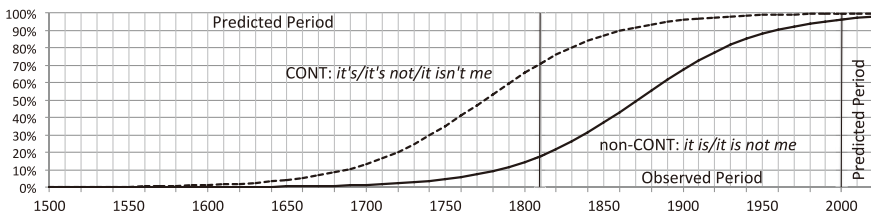

Figure 8. Shifts towards the use of *me* instead of *I* in CONT and non-CONT constructions

Predicted s-curves for the other accusative pronouns (*him/her/us/them*) are shown in Figure 9, with (a) non-CONT constructions on the left; (b) CONT constructions on the right. It shows that the curves for *him/her/us/them* remain at a level lower than those for *me* (●), and the presence of contraction encourages the shift towards ACC for all pronouns. The shift reached the level of near-completion (90%) in the 1960s when pronouns follow CONT constructions: *it's/it's not/it isn't*. This is true even for the pronoun *her* (▲), which shows the most-delayed shift of all. In contrast, the shift did not reach 90% in the same decade for *him/her/us/them* when they follow non-CONT constructions: *it is/it is not*.
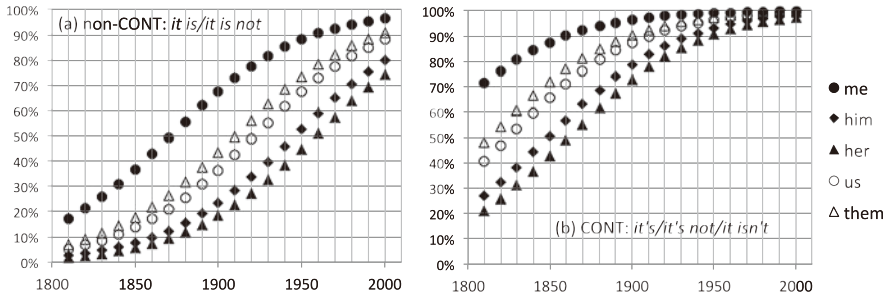
Figure 9. Predicted probabilities of ACC occurring in (a) non-CONT and (b) CONT constructions according to type of pronouns

4.4. Results and discussions

To sum up, what can be drawn from the above corpus research is that the case shift was nearly completed as early as the latter half of the 19th century for the first-person singular pronoun *I/me*, and the other pronouns followed *I/me*. The change progressed at a faster rate in sentences that contain a contraction (*it's/it's not/it isn't*) than in ones that do not (*it is/it is not*). In addition, the target construction [*it* + *is/was/were* + PERSONAL PRONOUN] occurred most frequently in the genre fiction (FIC) among the four genres examined here. These findings not only provide a piece of empirical support for the hypothesis that the shift towards ACC case is nearing completion, but also provide further details of the process of the syntactic change in question according to sentence construction (the presence/absence of contraction), type of pronouns, and register where the variation predominantly appears.

4.4.1. The relation of colloquialism and informality to the change

The advanced stage of change observed in contracted constructions (CONT) like *it's me* in the current research implies that the shift is more advanced in colloquial/informal contexts. A number of early grammarians admitted the common use of accusatives in spoken language in their age (e.g., Alford 1864: §192, Gowers 1962: 198, among others). Present-day researchers, including those who insist on the grammatical correctness of NOM cases, generally describe the use of ACC forms as predominant at least in informal style (Quirk et al. 1972: §4.112, Swan 1980: §135.1). According to Biber et al. (1999: §4.10.6.1), the change is most advanced in conversation to the degree that the diffusion of ACC forms is seen to be "nearly universal." The discussions above imply that the ACC forms spread from colloquial or informal domains.

The fact that the variation occurs predominantly in fiction (FIC) in COHA could be interpreted as support for the influence of colloquialism and informality on the diffusion of ACC. Colloquialism and informality are not only relevant to spoken language but also to certain genres in written language that abound in

quasi-conversation, contexts in which more than two participants are involved in the situation (e.g., fiction). To support this hypothesis, an additional investigation into the NOM-ACC alternation was conducted using BYU-BNC (British National Corpus). BNC, which was used as a synchronic corpus in this section, is a collection of 100 million words of written (90%) and spoken (10%) texts from British English in the late 20th century. The spoken component of this corpus consists of a wide range of genres (see Note 8) including more than four million words of "natural, spontaneous speech"[7] (Burnard ed. 2007: Section 1.5).

Criteria for retrieving data are the same as those stated in Section 4.1. As shown in Table 13, the shift towards ACC constructions exceeds 90% (303/313 = 0.968) in BNC overall. It is noted that the proportion of ACC usage surpasses 90% (208/214 = 0.972) in written fiction (W_FIC) and 100% if it is limited to the spoken component (S_CONV, S_OTHERS[8]). This result corresponds with what was observed in the research into COHA in the same decades (see Figure 2). It is notable that fiction (W_FIC) accounts for 68.4% of the total number of the collected tokens, although that genre comprises only 16.5% of the whole corpus. In addition, there is a similar tendency for conversation (S_CONV), which shows a larger proportion of the collected tokens (14.4%) than the share of that genre in the whole corpus (4.2%). That implies that fiction has something in common with conversation. It is most likely that conversation contains high proportions of colloquialism and informality; therefore, fiction is also most likely characterized as containing high proportions of colloquialism and informality.[9]

---

[7] The reason why BNC was chosen over American corpora (such as the Corpus of Contemporary American English (COCA), which a reviewer suggests that the author should analyze) is that BNC is considered a better option for highlighting the nature of fiction in terms of colloquialism and informality. The spoken component of BNC contains naturally-occurring spontaneous speech uttered by ordinary people in daily situations, while the spoken component of COCA consists of speech from a limited range of genres, i.e., TV and radio programs (for more details of the composition of COCA, see Davies 2009: 161).

[8] S_CONV and S_OTHERS correspond with the terms the *demographic* part and the *context-governed* part in Burnard (2007), or the *conversational* part and the *task-oriented* part in Leech et al. (2001: 2), respectively. S_OTHERS includes spoken data collected from: (i) educational and informative domains (e.g., lectures, news commentaries, etc.), (ii) business domains (e.g., company talks and interviews, trade union talks, etc.), (iii) public/institutional domains (e.g., political speeches, sermons, etc.), and (iv) leisure domains (e.g., speeches, sports commentaries, etc.). For a full list of domains included in the spoken component of the corpus, see http://www.natcorp.ox.ac.uk/docs/URG/ (accessed September 2020). The composition of BYU-BNC is available at https://www.english-corpora.org/bnc/ (accessed September 2020).

[9] This does not mean that how fictional characters speak completely matches actual language used in spontaneous speech. The reference to spontaneous speech here is to give an explanation of why the variation occurs predominantly in fiction based on common ground found between the two domains. An analysis of potential differences between the two domains is an interesting matter for discussion, but it will have to be left for further investigation.

In brief, the findings from the investigation into COHA and BYU-BNC clearly show that the *it is I/me* construction is genre-specific: the variation occurs mostly in the context which abounds in conversation/quasi-conversation, characterized as being colloquial and informal.

Table 13. Distribution of collected ACC and NOM tokens by genre compared with corpus size (BYU-BNC)

	Size of subcorpus (million words, %)		Number of collected tokens			
			ACC	NOM	ACC+NOM	
W_AC	15mw	15.9%	1	1	2	0.6%
W_FIC	16mw	**16.5%**	208	6	214	**68.4%**
W_MAG	7mw	7.5%	10		10	3.2%
W_MISC	21mw	21.6%	14	1	15	4.8%
W_NEWS	10mw	10.9%	14		14	4.5%
W_NON-AC	16mw	17.1%	5	2	7	2.2%
S_CONV	4mw	**4.2%**	45		45	**14.4%**
S_OTHERS	6mw	6.2%	6		6	1.9%
TOTAL	96mw	100.0%	303	10	313	100.0%

4.4.2. Popularity of *it is me* as opposed to *it is him/her/us/them*

The occurrence of *it is me* is popular in two ways. First, *me* shows the highest raw frequency among the pronouns under investigation (see Table 7). The popularity of the first-person singular pronoun *me* in this sense may stem from a close connection between the occurrence of *me* and the genre fiction, which is associated strongly with conversational register as Table 13 implies. As Smith (1906: 86) says, "the exigencies of colloquial English call for the first person singular far oftener than for any of the other forms" and presumably it is responsible for the higher frequency of *it is me* compared to *it is* plus any other pronouns in ACC case.

Second, the proportion of occurrence of ACC case to that of NOM case is highest for the first-person singular (Figure 9), i.e., the change has been pioneered by *I/me*. The advanced stage of the shift towards ACC constructions in the first-person singular would explain why the use of *him/her/us/them* as opposed to that of *me* has been felt to be "vulgar" (Onions 1904: §25).

The advanced stages of the change towards the use of *me* might give a useful insight into a mechanism for this syntactic change. One possible hypothesis is (i) that the change in case selection from NOM to ACC was initially brought to *I/me* by the phonological analogy of other NOM constructions such as *it is we/ye/he/she* (Jespersen 1894: §193); and (ii) that subsequently the shift towards ACC constructions spread into other personal pronouns due to the syntactic analogy of *it is me* (Sweet 1891: §1085), the ACC construction most frequently used in colloquial English.

4.4.3. Importance of usage-based grammar

The present corpus research, as a case study of American written English, has verified the well-known hypothesis that the case shift is nearing completion. The results from the corpus research above (Figure 9) can be compared with grammarians' statements particularly on American English discussed in Section 2.2.3 (Table 3). Mencken (1919: 222–223) stated in the 1910s that the use of ACC in the NOM position occurred in American English. Greet declared in Partridge (1963: 157) the widespread use of ACC during the 20th century by stating: "In America, *it's me* is acceptable colloquial English." Greet's statement corresponds with the data illustrated in Figure 8, where the shift towards *it's me* in COHA had been nearly completed by that time. It is interesting that a prescriptive attitude towards *it is I/me* was still preserved in written media as of 1933 (Evans & Evans 1957: 294, see Table 3), although the proportion of ACC usage in COHA had already exceeded 80% (for non-CONT) by that time (Figure 8). It is also worth noting that more than half of the members of the Usage Panel of *The American heritage dictionary of the English language* preferred *it is I* over *it is me* for formal writing as of 1969 (Morris ed. 1969: 810) while the proportion of ACC usage in COHA in the same decade had reached 90% for non-CONT and almost completed for CONT. In speech, in contrast, the use of ACC forms is termed acceptable by only 60% of the Panel. Even if the genre fiction in COHA is characterized as colloquial and informal as discussed in Section 4.4.1, the actual proportion of ACC variants in COHA (more than 90%) is still largely different from that shown in the Panel's level of acceptance of ACC usage in speech. The above discrepancy between people's perception of grammar and the actual usage of the language strongly suggests the importance of the pursuit of usage-based grammar.

5. Conclusions

The present research has provided empirical evidence for the diachronic shift towards the use of ACC forms instead of NOM forms after the verb *to be* from a viewpoint of corpus-based variationist linguistics (CVL). The shift towards ACC case in this construction has long been well known to linguists as grammar books and dictionaries show, and is considered to have been accepted as typical by the 20th century. The real-time data collected from the large-size diachronic corpus COHA has represented almost the entire process of the syntactic change and has provided empirical support for the claim that the change was nearing completion in the late 20th century in American written English. The additional investigation into the synchronic corpus BYU-BNC has shown that the shift was nearly complete in British English too in the late 20th century. The findings from the investigations of written register in COHA and BYU-BNC imply that the apparent connection between colloquialism/informality and the syntactic change in question, i.e., the change is advanced in spoken register or written register that abound in conversational settings. A CVL approach using the multivariate statistical model has succeeded in describing a more complex process of the change, in which the case shift was pioneered by the first-person singular *me* and contracted

predicate constructions (e.g., *it's/it's not*). The model also enabled one to predict the diffusion of ACC constructions, dating back to the reported beginning of the change (the 16th century).

The CVL methodologies in the current study helped obtain a deeper understanding of the well-known linguistic change. It has contributed to the development of the cross-disciplinary study of variation, combining advantages of variationist sociolinguistics with those of corpus-based linguistics. On the one hand, one of the contributions of the current study to corpus linguistics is the exact measurement of the growth of the ACC constructions as opposed to the frequency of their NOM counterparts. That allowed researchers to discern a change in a grammatical rule within the related linguistic environment defined by a variationist approach (*it is* + personal pronoun). On the other hand, one of the contributions of the current study to variationist sociolinguistics is that a model of the change is more accurate because predictions were made based on big real-time data. Note that predictions based on the contrasting method, apparent-time approach, could underestimate (Sankoff 2006: 115) or overestimate (Yokoyama & Sanada 2010, Kuya 2019: Ch. 8) a rate of change, since it uses differences in the age of language users to predict the speed of language change. Instead of using demographic factors like age, the present study attempted a real-time observation of grammar over a 200-year span of time in order to develop a model of language change. Large-scale real-time data made the statistical modeling of language change more rigorous and reliable.

Issues that need to be addressed in the future are two-fold. First, variationist investigations into the demographic aspects of the linguistic phenomenon will have to be combined with corpus linguistics. Wales (1996: 95) states that the NOM construction is perceived by younger generations to be "pedantic" or "stilted" whether the register is spoken or written, which implies that there is a difference in attitude towards ACC usage according to age; therefore this is a "generational change" (Labov 1994: 83). Early grammarians often mention, and might sometimes regret, that the use of ACC forms jumped across social class and reached educated speakers (e.g., Onions 1904: 34, §25), because that meant it had become well established in society. Visser (1970: 239, §264) points out a difference in the distribution of ACC forms in the 20th century according to three types of people: (i) those who use ACC constructions all the time ("illiterate people" in Visser's terminology); (ii) educated people who chose the case carefully according to register or style; (iii) educated people who exclusively use NOM constructions. A detailed analysis of the distribution of *it is I/me* in novels in the 19th century (Nakayama 2014) revealed the predominance of the use of ACC versions among those from lower social classes. Synthesis of the findings from the current study and an extended study into the above-mentioned demographic factors will add new insight into the theory on this language change.

Second, to obtain extensive knowledge of the NOM-ACC alternation phenomenon, consideration of other related constructions is required. They include constructions with (i) a relative clause that may cause relative attraction (e.g., *It is*

he/him that/who is looking for you) and (ii) a conjunction (e.g., *She is taller than I/me, He is as tall as I/me*).[10] As the data provided by Dekeyser (1975) and Biber et al. (1999) suggest, it is possible that the shift towards ACC case in constructions followed by a relative clause like *it was I/me who…* could have progressed at a slower pace due to syntactic restrictions imposed by relative attraction. The hypothesis that such syntactic restrictions slow the speed of change should be investigated to make a further generalization of the process of this syntactic change.

The scrutiny of the social factors and linguistic environments mentioned above will contribute to the extended description of usage-based grammar as opposed to prescriptive grammar of the English language.

References

Aitchison, Jean (1981) *Language change: Progress or decay?* (Fontana Linguistics). London: Fontana.

Alford, Henry (1864) *The queen's English: Stray notes on speaking and spelling*. London: Strahan. https://archive.org/details/queensenglishman00alfo/ [accessed November 2019].

Alford, Henry (1865) *A plea for the queen's English: Stray notes on speaking and spelling*. Second edition. London: Alexander Strahan. https://babel.hathitrust.org/cgi/pt?id=hvd.hwk7nv [accessed November 2019].

Barber, Charles (1976) *Early Modern English*. London: André Deutsch.

Bauer, Laurie (1998) You shouldn't say 'It is me' because 'me' is accusative. In: Laurie Bauer and Peter Trudgill (eds.) *Language myths*, 132–138. London: Penguin Books.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan (1999) *Longman grammar of spoken and written English*. Harlow: Longman.

Brook, G. L. (1958) *A history of the English language*. London: André Deutsch.

Brown, Goold (1827) *The institutes of English grammar, methodically arranged*. Third edition. New York: Samuel Wood & Sons. https://babel.hathitrust.org/cgi/pt?id=wu.89099896441 [accessed November 2019].

Brown, Goold (1856) *The institutes of English grammar, methodically arranged*. New stereotype edition. New York: Samuel S. & William Wood. http://dbooks.bodleian.ox.ac.uk/books/PDFs/600071475.pdf [accessed November 2019].

Burnard, Lou (ed.) (2007) *Reference guide for the British National Corpus* (XML edition). Published for the British National Corpus Consortium by the Research Technologies Service at Oxford University Computing Services. http://www.natcorp.ox.ac.uk/docs/URG/ [accessed September 2020].

Bybee, Joan (2006) From usage to grammar: The mind's response to repetition. *Language* 82(4): 711–733.

Celce-Murcia, Marianne and Diane Larsen-Freeman (1983) *The grammar book: An ESL/EFL teacher's course*. Rowley: Newbury House Publishers.

Cobbett, William (1819) *A grammar of the English language, in a series of letters*. London: Printed for Thomas Dolby. https://babel.hathitrust.org/cgi/pt?id=njp.32101063604878 [accessed November 2019].

Costello, Robert B. (ed.) (1991) *Random House Webster's college dictionary*. New York: Ran-

---

[10]Furthermore, the case shift appears to be progressing in the opposite direction (from ACC to NOM) in the following examples: (iii) *Between you and me/I…*; (iv) *He is the man whom/who you are looking for*.

dom House.

Curme, George Oliver (1931) *Syntax*. Boston: Heath (Reprinted as the Asian ed. by Maruzen, Tokyo, 1959).

Davies, Mark (2009) The 385+ million word *Corpus of Contemporary American English* (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics* 14(2): 159–190.

Davies, Mark (2012) Expanding horizons in historical linguistics with the 400-million word Corpus of Historical American English. *Corpora* 7(2): 121–157.

Dekeyser, Xavier (1975) *Number and case relations in 19th century British English: A comparative study of grammar and usage*. Antwerpen: Uitgeverij de Nederlandsche Boekhandel.

Erdmann, Peter (1978) *It's I*, *it's me*: A case for syntax. *Studia Anglica Posnaniensia* 10: 67–80. http://hdl.handle.net/10593/10713 [Accessed September 2020].

Evans, Bergen and Cornelia Evans (1957) *A dictionary of contemporary American usage*. New York: Random House.

Fowler, Francis George and Henry Watson Fowler (1924) *The pocket Oxford dictionary of current English*. Oxford: Clarendon Press.

Fowler, Henry Watson and Francis George Fowler (1906) *The king's English*. Second edition. Oxford: Clarendon Press. https://archive.org/details/kingsenglish00fowlrich/ [accessed November 2019].

Görlach, Manfred (1999) *English in nineteenth-century England: An introduction*. Cambridge: Cambridge University Press.

Gowers, Ernest (1962) *The complete plain words* (Pelican Books, A554). Harmondsworth: Penguin Books.

Han, Weifeng, Antti Arppe and John Newman (2013) Topic marking in a Shanghainese corpus: From observation to prediction. *Corpus Linguistics and Linguistic Theory*, 1–29. DOI: 10.1515/cllt-2013-0014.

Hornby, Albert Sydney (1974) *Oxford advanced learner's dictionary of current English*. Third edition. London: Oxford University Press.

Hosmer., David W., Stanley Lemeshow, and Rodney X. Sturdivant (2013) *Applied Logistic Regression*. Third edition. Hoboken, New Jersey: Wiley.

Huddleston, Rodney and Geoffrey K. Pullum (2002) *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Jespersen, Otto (1894) *Progress in language with special reference to English*. London: Swan Sonnenschein & Co. https://archive.org/details/cu31924026448203 [accessed September 2019].

Jespersen, Otto (1933) *Essentials of English grammar*. London: Allen & Unwin.

Jewell, Elizabeth J. and Frank Abate (eds.) (2001) *The new Oxford American dictionary*. New York: Oxford University Press.

Kellner, Leon (1892) *Historical outlines of English syntax*. London: Macmillan. https://archive.org/details/historicaloutlin00kelliala [accessed November 2019].

Kisbye, Torben (1972) *An historical outline of English syntax, Part II*. Aarhus: Akademisk Boghandel.

Kuya, Aimi (2019) *The diffusion of Western loanwords in contemporary Japanese: A variationist approach*. Tokyo: Hituzi Syobo.

Labov, William (1972) *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William (1994) *Principles of linguistic change, Vol.1: Internal factors*. Oxford: Blackwell.

Latham, Robert Gordon (1862) *The English language*. Fifth edition. London: Walton and Maberly. https://archive.org/details/englishlanguage00lath/ [accessed November 2019].

Leech, Geoffrey and Jan Svartvik (1975) *A communicative grammar of English*. London: Longman.

Leech, Geoffrey, Paul Rayson, and Andrew Wilson (2001) *Word frequencies in written and spoken English: Based on the British National Corpus*. Harlow: Longman.

Lounsbury, Thomas Raynesford (1894) *History of the English language*. Revised and enlarged edition. New York: Henry Holt and Company. https://archive.org/details/historyofenglish00loun [accessed September 2019].

Lowth, Robert (1762) *A short introduction to English grammar: With critical notes*. London: Printed by J. Hughs; for A. Millar; and R. and J. Dodsley (Reprinted by Scholar Press, Menston, 1967).

Lyons, John (1968) *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press.

Mason, Charles Peter (1879) *English grammar, including the principles of grammatical analysis*. Enlarged and thoroughly revised, 27th edition. Toronto: Adam Miller & Co. https://archive.org/details/englishgra79west00masouoft [accessed September 2019].

Mencken, Henry Louis (1919) *The American language: A preliminary inquiry into the development of English in the United States*. New York: Alfred A. Knopf. https://archive.org/details/americanlanguage00mencuoft [accessed September 2019].

Miller, Jim (2011) *A critical introduction to syntax*. London: Continuum.

Minami, Masahiko (2009) *Gengo to bunka* [*Language and culture*]. Tokyo: Kurosio.

Morris, William (ed.) (1969) *The American heritage dictionary of the English language*. Boston: American Heritage Publishing.

Murray, James Augustus Henry (ed.) (1933) *The Oxford English dictionary* (being a corrected re-issue with an introduction, supplement, and bibliography of *A new English dictionary on historical principles*), *Vol. VI: L–M*. Oxford: Clarendon Press.

Mustanoja, Tauno F. (1960) *A Middle English syntax, Part 1: Parts of speech*. Helsinki: Société Néophilologique.

Nakayama, Masami (2014) *It is I* vs. *It is me* in 19th-century English novels. *Studies in Modern English* 30: 1–23.

Onions, Charles Talbut (1904) *An advanced English syntax: Based on the principles and requirements of the grammatical society*. London: Swan Sonnenschein & Co. https://archive.org/details/anadvancedengli00oniogoog [accessed November 2019].

Partridge, Eric (1963) *Usage and abusage: A guide to good English* (Penguin Reference Books, R24). Fifth edition. Harmondsworth: Penguin Books.

Pearsall, Judy (ed.) (1998) *The new Oxford dictionary of English*. Oxford: Clarendon Press.

Potter, Simeon (1975) *Changing English*. Second rev. edition. London: André Deutsch.

Poutsma, Hendrik (1916) *A grammar of Late Modern English, Part II: The parts of speech, Section IB, Pronouns and numerals*. Groningen: Noordhoff.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik (1972) *A grammar of contemporary English*. London: Longman.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik (1985) *A comprehensive grammar of the English language*. London: Longman.

Sankoff, Gillian (2006) Age: Apparent time and real time. In: Keith Brown (ed.) *Encyclopedia of language and linguistics*, *Vol. 1*. Second edition, 110–116. Amsterdam: Elsevier.

Sapir, Edward (1921) *Language: An introduction to the study of speech*. New York: Harcourt, Brace and Company. https://archive.org/details/languageanintrod00sapi [accessed Novem-

ber 2019].

Smith, Charles Alphonso (1906) *Studies in English syntax*. Boston: Ginn & Company. https://archive.org/details/studiesinenglis01smitgoog [accessed November 2019].

Swan, Michael (1980) *Practical English usage*. Oxford: Oxford University Press.

Sweet, Henry (1891) *A new English grammar: Logical and historical, Part I*. Oxford: Clarendon Press.

Szmrecsanyi, Benedikt (2017) Variationist sociolinguistics and corpus-based variationist linguistics: Overlap and cross-pollination potential. *Canadian Journal of Linguistics/ Revue canadienne de linguistique* 62(4): 685–701.

Vallins, George Henry (1966) *The pattern of English*. Rev. edition. London: André Deutsch.

Visser, Frederik Theodoor (1970) *An historical syntax of the English language, Part I* (reprinted with corrections). Leiden: Brill.

Wales, Katie (1996) *Personal pronouns in present-day English*. Cambridge: Cambridge University Press.

Weiner, Edmund S.C. and Joyce M. Hawkins (1985) *The Oxford guide to the English language* (Oxford Paperback Reference). Oxford: Oxford University Press.

Yokoyama, Shoichi and Haruko Sanada (2010) Gengo no shoogai shuutoku moderu ni yoru kyootsuugoka yosoku [Predictions of dialect standardization by the "life-long assimilation of language change" model]. *Nihongo no Kenkyu* 6(2): 31–45.

Zandvoort, Reinard Willem (1965) *A handbook of English grammar*. Third edition. London: Longmans, Green & Co.

Online Corpus & Dictionary

BYU Corpora. https://corpus.byu.edu/ [accessed September 2019].

OED Online. https://www.oed.com [accessed September 2019].

Author's contact information: [Received 30 May 2020;
Fukuoka Jo Gakuin University   Accepted 23 October 2020]
3-42-1 Osa, Minami-ku, Fukuoka, 811-1313 JAPAN
e-mail: kuya@fukujo.ac.jp

【要　旨】

It is I／*It is me* 構文のコーパス基盤変異理論的考察
——ゆれがほぼ完了した統語変化のCOHAを活用した実時間調査——

久屋　愛実

福岡女学院大学

　本稿は，ほぼ完了したとされる「*it is* 人称代名詞」構文における主格から目的格への交替を，コーパス基盤変異理論的視点（CVL）から捉えた実証的研究である。先行研究は新変異形がすでに一般に普及していることを示唆するが，この変化の全容を捉えた実証データは管見の限り存在しない。そこで，通時的コーパス COHA の実時間データを計量的に分析し，変化のほぼ全過程を予測する多変量モデルを提示した。その結果，変化は（1）一人称単数形（*I/me*）で最も早く進行したこと，（2）*I/me* では 20 世紀後半までにほぼ完了したこと，（3）非縮約構文（*it is*）よりも縮約構文（*it's, it isn't* 等）の環境下で先に拡大したことが判明した。本研究は，コーパス言語学と変異研究の利点を統合させた分野横断的アプローチにより，長年にわたり議論されてきた統語変化の過程を実証的，包括的かつ詳細に記述することに成功した点で，言語変異・変化研究に資する。