REFLEXES OF THE TRANSITIVE BE PERFECT IN CANADA
AND IN THE US: A COMPARATIVE CORPUS STUDY

Yuri YERASTOV

Fort Hays State University*
y_yerastov@fhsu.edu

Abstract

The present paper examines the geographical distribution of three reflexes of a transitive be perfect in North America: done, finished, started (e.g. I'm done my dinner). This paper is a corpus study of Canadian Newsstand Complete and Newspaper Source Plus — commercial databases of regional and national mass media in Canada and the United States, respectively. A total of 1217 tokens have been found in Canadian sources, and none in the US ones. Constructional tokens in Canada have been found to be proportionately distributed across provinces and municipalities. Based on these findings, it is argued that the done my dinner construction is a feature that sets Canadian English (as a theoretical abstraction) apart from American English. It is emphasized, though, that the construction has been attested in some dialectal pockets in the US such as Vermont. The geographical restriction of the construction to Canada and selected dialectal areas in the US is hypothesized to have arisen from a Scottish founder effect; evidence from fictional literature and migration studies is presented as support. It is further argued that the construction is a retention in North American English rather than innovation because its earliest occurrence dates back to the middle of the 19th century.

Keywords

transitive be perfect, geographical distribution, dialects, Canadian English, Scots, I am done my dinner

* 600 Park Street, Department of English, Hays, KS 67601-4099, USA.
REFLEJOS DEL TRANSITIVO PERFECTO DE SER EN CANADÁ Y USA:
UN ESTUDIO COMPARATIVO DE CORPUS

Resumen
Este artículo analiza la distribución geográfica de los tres reflejos del transitivo perfecto de ser en América del Norte: hecho, acabado, iniciado (por ejemplo, he terminado mi cena). Este trabajo es un estudio de corpus del Canadian Newsstand Complete y el Newspaper Source Plus — bases de datos comerciales de los medios de comunicación regionales y nacionales de Canadá y Estados Unidos, respectivamente. Se han encontrado 1217 registros en las fuentes canadienses, y ninguno en la de los Estados Unidos. Se han encontrado registros de construcciones en Canadá para ser distribuidos proporcionalmente entre provincias y municipios. Sobre la base de estos hallazgos, se discute que la construcción done my dinner es una característica que define el inglés canadiense (como abstracción teórica), diferente del inglés americano. Se hace hincapié, sin embargo, que la construcción se ha registrado en algunos puntos dialectales de los EE.UU., como Vermont. Se hipotetiza que la restricción geográfica de la construcción en Canadá y en las áreas dialectales seleccionadas en los EE.UU. puede haber surgido del efecto fundacional del escocés, según parecen indicar la literatura de ficción y los estudios de migración. Se argumenta además que esta construcción es una retención en el inglés de América del Norte más que una innovación, ya que su primera aparición se remonta a mediados del siglo XIX.

Palabras clave
transitivo perfecto de ser, distribución geográfica, dialectos, inglés canadiense, escocés, I am done my dinner

1. Introduction

One distinct feature of Canadian English that sets it apart from American English has, until recently, eluded dialectological and theoretical treatments — namely, the absence of a preposition after done and finished in exemplars such as:

(1) a. And the kids are all either finished school or finishing off university.
   b. ...I would come out and play ball when we were finished our work.
   c. Ontario Provincial Police – when they were finished their investigation – they...
   d. I’m almost finished the fifth of those ten...
   e. Are you finished your question?
   f. ...because we’ll be finished the application before...

Strathy Corpus of Canadian English
These exemplars contrast with their Standard English equivalents *I am done with dinner* and *I am finished with my homework*. The underlying grammatical schema is not limited to *done* and *finished* only; in some Canadian dialects, it may also be extended to *started*, e.g. *I am started my homework* (Yerastov 2010, 2011, 2012). In the present paper, I will use the shorthand notation *[be done NP]* to refer collectively to the constructional schema *[be {done, finished, started} NP]*.

Prior literature has reported on various reflexes of the transitive *be* perfect in Bungi English in the Canadian prairies (Gold 2007), Lumbee English in North Carolina (Wolf ram 1996), and dialects spoken in Southern Atlantic states and Pennsylvania (Atwood 1953). It has also been experimentally shown that the construction *[be done NP]* is acceptable to speakers of Canadian English in Alberta, and is likely to be a special case of the transitive *be* perfect (Yerastov 2012).

The comparative corpus study, reported on in this paper, tracks the geographical distribution of the construction¹ in Canada and the US, using data from mass media databases. The results of the corpus study show that *[be done NP]* is widespread in Canadian English, and marginal in American English. The results further show that the construction is proportionately distributed across Canadian provinces and municipalities, with *done* and *finished* preferred as participles in the construction. I motivate the geographical distribution of the construction with a pattern of historical settlements, and draw connections to evidence of *[be done NP]* in historical corpora and fictional literature. In doing so, I reinforce the hypothesis of Scottish etymology of *[be done NP]* proposed in prior work (Yerastov 2010, 2011, 2012)

With a corpus focus, the present paper complements etymological conclusions drawn from experimental data (2012) but differs from the experimental work in having its main focus on the geography of the construction rather than its morphosyntax.

¹ Henceforth, I use *[be done NP]* to reductively refer to *[be {done, finished, started} NP]* unless greater exactitude is needed.
Unlike previous corpus-based work on \([be\ done\ NP]\) (Yerastov 2010, 2011), the present paper is based on a substantially larger, systematically expanded sample of data.

The paper is organized as follows. In §2, I review the literature on transitive be perfects in English, suggesting a connection between various dialectal attestations of the transitive be perfect and \([be\ done\ NP]\). In §3, I discuss the sources of my data, the search procedure, and a control condition for my study. In §4, I present my findings, demonstrating the prevalence of the construction in Canada and its marginality in the US. In §5, I highlight the most important findings, and situate them in relation to some of the literary, historical, and dialectal facts; I tentatively propose a Scottish etymology for the construction. In §6, I conclude with a list of unresolved problems, and directions for future research.

2. Literature review

2.1 Periphrastic perfects and split auxiliary systems

Perfect constructions in Indo-European languages normally take the form of verbal periphrasis in which an auxiliary — either be or have — combines with a past participle form. The emergence of periphrastic perfect constructions follows well-established grammaticalization paths (Bybee et al. 1994); the input to grammaticalization comes from stative constructions for be perfects and possessive constructions for have perfects. Because possession inherently requires a direct object, have perfects, after grammatical reanalysis, inherit that object. On the other hand, statives do not inherently require a direct object; periphrastic perfect structures that emerge from stative sources do no have a syntactic position for direct objects. Therefore have perfects are inherently transitive, while be perfects inherently intransitive. It is well known that the two periphrastic perfect schemas compete with each other for membership of past participles. In some languages, such competition results in split auxiliary systems, as for example in French or German. In others, the have schema takes over the perfect domain, e.g. English or Pennsylvania German (Heine & Kuteva 2005: 140-41).
Reverse developments, in which the auxiliary *be* takes over the entire perfect domain, appear to be possible as well, but they are not well – if at all – documented diachronically, being inferred from synchronic states. For example, within the Indo-European family, very few dialects/languages are known to have a fully productive transitive *be* perfect: insular dialects of Scots – Shetlandic (Robertson & Graham 1991) and Orcadian (Flaws & Lamb 1996), Lumbee (Wolfram 1996) and Bungee (Gold 2007) English in North America, and dialects of Italian (Bentley & Eythórrsson 2003: 451); but their origin remains elusive beyond speculations of cross-dialectal transmission or reanalysis of *have* as *be*. Either way, dialectally in Present Day English (PDE) we observe split auxiliary systems in which the auxiliaries *be* and *have* alternate in selecting for a limited set of transitive participles, as is seen today in Canadian English, as well as several dialects of American English.

On the synchronic level, *be~have* auxiliary splits have been approached from both syntactic and semantic perspectives or a combination thereof, but crucially they have fallen short of tackling alternations involving the transitive *be* perfect (see Yerastov (2105) for a critique of these approaches) and have little light to shed on the theoretical understanding of [*be done NP*].

### 2.2 Etymology of [*be done NP*]

Since the emergence of [*be done NP*] is not historically documented, we can use the construction’s synchronic clues as a window onto its past. The construction yields resultative interpretations, as would be expected from a perfect construction in English (Yerastov 2011, 2012); for example, a speaker of Canadian English saying *I’m done the dishes* profiles the result of the event, which persists at speech and reference time. Resultative interpretation is what [*be done NP*] shares in common with its better known intransitive relative [*be done (with NP)*]. In fact, Yerastov (2012) notes that in PDE, the resultative meaning of the intransitive *I am done/ finished* is odd language-internally, because other dynamic verbs are incapable of such behavior. Formally similar structures such as *I am ruined, I am screwed, I am messed up* can only be understood as stative passives, in which the subject is a patient or experiencer, but never an agent. Exemplars
such as *I am done/ finished*, on the other hand, uniquely combine the two interpretations:

(3) I am finished

  'I have finished' (resultative)
  'My life/ career is finished' (stative passive)

Because of the similarity in the resultative semantics of the [*be done (with NP)*] and [*be done NP*] constructions, we can hypothesize that they are genetically related, and look at the history of the intransitive participles *done* and *finished* in search of insights into the common origin of the two constructions.

Is it possible that the co-existence of both resultative and stative passive interpretations in PDE, as shown in (3), represents a synchronic layering resulting from recent grammaticalization? After all, it is diachronically typical for resultatives to develop from passive statives (Bybee *et al.* 1994). However, a brief examination of historical data from the Oxford English Dictionary (OED) suggests an unlikelihood of such a scenario. OED reports the first resultative uses of the intransitive participles *done* and *finished* to have occurred rather late — at the end of the 18th century: *I was done with love for ever* (1766); *One farther favor and I am done* (1771); *The rogue is pressing and I must be done* (1776), followed by consistent attestation in the 18th century. The first resultative use of [*be finished*] is attested as late as in the 20th century: *How often have you told me that you are finished with all women!* (1939). It is unlikely that [*be {done, finished}*] grammaticalized from a stative passive into a resultative around the 18th century or later, because within this timeframe the *be* perfect paradigm was unproductive and likely incapable to attract new members.

In fact, such grammaticalization would not have been likely at any point in the history of mainstream British English. According to OED, the verbs *finish* and *start* are not autochthonous, Old English verbs. Derived from Old French, the verb *finish* is first attested in Middle English in 1375. While the origin of *start* is questionable, OED suggests that it is likely derived from a Scandinavian source and documents its first occurrence in 1000. A relatively late entrance of these verbs into the English lexicon.
suggests they were unlikely to develop resultative uses, since at the time of their entrance the intransitive be perfect paradigm was already starting to lose its schematicity and was unlikely to be extended to new members, either.

More to the point, OED characterizes this construction as chiefly Irish, Scottish, US, and dialectal, which suggests that it arose outside of mainstream British English, a conclusion that can be hypothetically extrapolated to the transitive [be done NP] because of synchronic similarity between the two constructions.

In what follows I review literature that elucidates the possibility of [be done NP] arising as a result of contact with Gaelic and Scots since these varieties have a be perfect.

2.2.1 Gaelic

Owing to its Gaelic substratum, Irish English has the after perfect, which combines with the auxiliary be. A reflex of that perfect is found in dialects of Irish English including those in eastern Canada:

(4) You’re after ruinin’ me
   ‘You have ruined me’
   (Filppula 1999: 90)

This after perfect is formally and semantically paralleled by an immediate perfect construction in Irish:

(5) Tá said tar éis teach a thógáil
    Is they after house build –VN
    ‘They are after building a house’.
    ‘They have just built a house.’
    (Hickey 2007: 149)
The presence of the construction [be after V-ing] in Gaelic may certainly have reinforced the usage of be as perfect auxiliary in the construction [be V-en NP], but a direct transformation of one into the other seems to be implausible due to a wide formal gap. Crucially, Siemund (2003) finds superstrate accounts of perfect constructions in Irish English to be more cross-linguistically plausible.

2.2.2 Scots and Irish English

A more likely candidate for the etymological source of [be done NP] is insular Scots. There are abundant attestations of the transitive be perfect in Shetlandic and Orcadian, two varieties of insular Scots. Millar (2007: 75) reports that “the most striking structural feature of Shetlandic is the use of be as an auxiliary verb in active perfective construction with all types of verbs”. Reference works are also unanimous in recognizing the transitive be perfect in Shetland and Orkney. In an introduction to Shetland grammar, Robertson & Graham (1991) cite a number of instances of the transitive be perfect:

(6) a. Fifty voars I’m dell’d an set da tatties
    Fifty spring I’m sorted and planted the potatoes
b. When A’m feenished yun A’ll be dön a göd day’s wark
    When I’m finished that I’ll be done a good day’s work
c. Ye never did ony ill an’ noo ye’re dune me muckle guid
    You never did any ill and now you’re done me much good
d. I’m read my Bible.
   (Robertson & Graham 1952: 11)

In a grammatical overview of Orcadian, Flaws & Lamb (1996) state that the auxiliary be is used instead of English have:

(7) a. Ah’m meed the dinner
    I’m made the dinner
b. Wir biggid the stack
   We’re built the stack

c. Thoo’ll be gotten a fair price for thee kye
   You’ll be gotten a fair price for your cattle
   (Flaws & Lamb 1996: X)

These attestations may be independently confirmed with data from Shetlandic sources in the Scottish Corpus of Texts and Speech (SCOTS):

(8) a. heidmaster realised we were done wur [our] bit
    b. A’m funn dem!
       I am found them
    c. A’m read dem ower
       I am read them over
       Scottish Corpus of Text and Speech

With regard to Shetlandic, Pavlenko (1997) hypothesizes that the transitive be perfect has arisen from the reanalysis of homophonous third-person singular forms of the auxiliaries have and be and subsequent paradigm leveling in favor of be. These processes were likely spurred by contact between Scots and Norn that dates back to the 15th century.

The productivity of the transitive be perfect in insular Scots and its survival to the present day leave open the possibility that there have been multiple opportunities for some of its exemplars to migrate to North America.

The restriction of the transitive be perfect to a set of participles done, finished, started, found in the present study, resembles the situation in non-insular Scots, and Irish English. One particular exemplar of the transitive be perfect, I am finished it, is so prominent in Scots that Trudgill & Hannah (1982: 88) view it as a feature of Scottish English, as opposed to Standard English exemplars such as I am finished and I have finished it. An examination of the Scottish Corpus of Texts and Speech confirms the presence of finished exemplars in non-insular Scottish sources:
a. We're nearly finished this ain't we.
b. We are finished this ones and this ones and this ones, Mammy.
c. I'm finished something.
d. You can get ain when we're finished our tape.
e. We're just about finished the sewing.

Scottish Corpus of Texts and Speech

At the same time, started exemplars have been attested by Caroline Macafee as a non-standard feature of mainland Scots.³

(10) when I was just started school in the babies class, ...

A similar situation obtains in Irish English. Hickey (2007: 178) reports that speakers of Irish English in such geographically separated areas as Derry, Kerry, Offaly, and Monaghan counties accept the exemplar They’re finished the work now – with a mean of 85%. While a search of the Corpus of Irish English does not yield any results, independent confirmation may be found in the Irish Internet domain (ie.), which contains all three participles finished, started, done.

(11) a. It is important that I am finished my morning routines by ten o’clock [...].
   b. I intend to continue acting classes well after I am finished the bronze award.
   c. Now I am done my song, boys, but yet don’t go away [...]. (Waterford Songs)
   d. Yes, she is done a great job [...].
   e. I am started the gym on Monday and plan on going 4 times a week.

Irish Internet domain (Google search)

The similarity of entrenchment patterns of [be done NP] in North America, Scotland, and Ireland suggests a genetic relationship.

³ https://docs.google.com/file/d/0BzVAfXkKg9UJVZ2dwnERCBUw7SGc/edit
2.2.3 Reflexes in North America

Some exemplars of the transitive *be* perfect have been attested in North America (Atwood 1953, Wolfram 1996, and Gold 2007). Atwood (1953: 26-27) notices both transitive and intransitive instantiations of the *be* perfect in Southern Atlantic states – (9a) & (9b), and Pennsylvania – (9b), making a vague connection to British dialects.

(12) a. I am heard it
   b. I am been thinking

More recent reports document similar occurrences in Bungi English in the Canadian prairies (Gold 2007), and in Lumbee English in Robeson county, North Carolina (Wolfram 1996).

(13) a. I am not got the horse tied upset the Hotel
   b. Aw Willie, I am just slocked [‘extinguished’] it the light
   c. he’s bin so greedy
   (Gold 2007)

(14) a. If I’m got a dollar I’m got it.
   b. I says, I’m Indian, I says, I’m been nothing, I says, but a Indian, I says here.
   (Wolfram 1996)

Neither Gold (2007) nor Wolfram (1996) provide much descriptive evidence of the productivity of the transitive *be* perfect. Gold acknowledges that in her study of a corpus of Bungi data there are only a few unambiguous examples of the transitive *be* perfect where the auxiliary is unmistakably *be*, as in (10a) & (10b), as opposed to (10c), where ’s is ambiguously homophonous and may represent either *be* or *have*. Likewise, Wolfram’s examples are limited in productivity to the verbs *got* and *been*.

The origin of likely reflexes in North America is unsettled, as there is no direct, conclusive evidence. Wolfram (1996) seems to appeal more to language-internal resources as an explanation, while alluding, in passing, to the possibility of Scottish and
Scotch-Irish influence on Lumbee English in general. Gold (2007) points out that speakers of Bungi English in the Canadian prairies descended from employees of the Hudson Bay Company, who were systematically recruited, across several generations, in the Highlands of Scotland and the Orkney islands. She concludes that a Scottish etymology is more likely than other explanations such as vernacular universals (Chambers 2003). While neither account is conclusive in isolation, both of them point to a Scottish etymology for the transitive be perfect in North American English. If this conclusion is true, we should expect that the geographic distribution of the construction [be done NP] in North America will reflect Scottish settlement patterns.

An experimental study (Yerastov 2012) has demonstrated that the construction [be done NP] occurs in Canada, an area with a documented Scottish founder effect (Dollinger 2008; Benett 2003; Bumstead 1981). While most participants in the study came from Calgary and the province of Alberta, Yerastov (2012) argues that the results are generalizable to Canadian English as a whole. Other experimental studies confirm the absence of the construction in Illinois, but document its occurrence in Vermont (Yerastov 2010), another area where a Scottish founder effect has been documented (Shields 1996). Yerastov (2012, 2010) further demonstrates that the construction is not an idiom, but rather partially schematic. Thus, the subject slot is restricted to animate referents, the participle slot favors three items only, and the direct object slot tends to be marked for definiteness. However, the direct object slot is open-ended and variable. Importantly, Yerastov (2012) finds that other participles such as made, heard, read received marginal ratings of acceptability among his Canadian participants, a fact that suggests that there is a gradient continuum between [be {done, finished, started} NP] and other exemplars of the transitive be perfect, such as I am heard deer in the bush.

While experimental studies offer the advantage of testing the morphosyntactic limits of the construction, these studies have an inherent drawback in that they are skewed to the locale from which the participants are drawn, not to mention the observer’s paradox (Labov 1972). It is understandably difficult to recruit a participant sample representative of entire North America. Even though the participant samples in Yerastov (2012, 2010) included a few consultants from areas other than Alberta, Illinois and Vermont, one needs to independently confirm the generalizability of such findings.
to the entire countries of Canada and the United States. The present study attempts to
do so with a corpus study of Canadian and American English. Unlike previous work, the
focus of the present study is on geographical distribution; no attempt is made here to
refine previous morphosyntactic and semantic analyses.

3. Methodology

3.1 Goal

The goal of the present investigation is to track the geographical distribution of
the construction \([\text{be done NP}]\) in the US and Canada.

3.2 Hypothesis

I hypothesize that \([\text{be done NP}]\) is prevalent in Canada and marginal in the US. My
hypothesis is informed by the distribution of the transitive \(\text{be perfect}\) across Scottish/
English dialects (Millar 2007; Pavlenko 1997; Flaws 1996), occurrences of \([\text{be done NP}]\) in
Canada (Yerastov 2012), and a Scottish founder effect on Canadian English (Dollinger

3.3 Sources of data

While there have been successful corpus studies of English dialects (e.g. Schneider
& Montgomery 2001; McCafferty 2003; Van Herk & Walker 2005), there are still some
challenges one needs to overcome in such studies. First, there are very few dialectal
corpora in English, with the notable exception of the Helsinki Dialect Corpus, the
Freiburg English Dialect Corpus, and dialectal annotations in the British National Corpus.
Second, dialectal features are often too scarce in corpus data; a feature has to be
relatively frequent, \(\geq 1000\) tokens per running text (Szmrecsany 2008), to afford
meaningful generalizations.
To be sure, these challenges came up in my study of the construction \([be\ done\ NP]\). My searches for tokens of \([be\ done\ NP]\ in well-established linguistic corpora returned marginal results, which did not correspond to the robustness of the construction, which I informally observed in Canadian and Vermont English. Thus the Corpora of Contemporary and Historical American English returned only 6 tokens, the Strathy Corpus of Canadian English – 8 tokens, and the Bank of Canadian English – 2 tokens (see Table 10 below). While these results provided valuable attestations of the construction (some of which will be discussed in § 5), they were not statistically meaningful for tracking the geographical distribution of the construction. Therefore I decided to turn to mass media databases as the source of corpus data for the present study.

My search of mass media sources was motivated by the assumption that tokens of \([be\ done\ NP]\ would occur, with sufficient frequency, in regional and local newspapers. Accordingly, I chose the commercial databases Canadian Newsstand Complete, and Newspaper Source Plus. Canadian Newsstand is a full-text database of 355 Canadian newspapers. Newspaper Source is a full-text database of 729 US newspapers (along with international periodicals as well, which were excluded from search results). The comparative descriptors of the two databases are summarized in Table 1; a complete list of news media sources in these databases can be obtained from the websites of their vendors: ProQuest and EBSCOHost.

<table>
<thead>
<tr>
<th>database</th>
<th>country</th>
<th>years covered</th>
<th>no. periodicals</th>
<th>no. days of coverage$^4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newspaper Source</td>
<td>USA</td>
<td>1990-2014</td>
<td>729</td>
<td>1,525,615</td>
</tr>
<tr>
<td>Canadian Newsstand</td>
<td>Canada</td>
<td>1977-2014</td>
<td>355</td>
<td>993,775</td>
</tr>
</tbody>
</table>

Table 1. Comparison of databases

A particular challenge that a corpus linguist faces is the absence of information on the total number of words in news media databases. Because these databases are updated with new editions, the providers do not index the databases for word count. A corpus linguist must therefore find alternative ways of quantifying the size of dynamic

$^4$ This figure is calculated by summing up the totals of days of coverage for each periodical in the database. Overlapping coverage is not subsumed but added up.
corpora, based on descriptors provided by database vendors, in order to have a meaningful yardstick against which to interpret and normalize search results.

One straightforward solution might be to measure the frequency of a linguistic feature against the number of periodicals in a database. This solution is not without problems because periodicals differ in how far they go back in time. A more accurate solution might be to count the number of days of coverage for each periodical in a database and sum up the results. Thus, one could quantify a universe of discourse covered by database material in terms of time rather than word count. Adopting this approach, we can sum up the days of coverage across all periodicals within each database (Table 1) and calculate the normalization factor for Canadian search results: $1,525,615 : 993,775 = 1.5$. Of course, this course-grained approach is not without flaws either; for one, the length of a periodical varies from week to week and from publisher to publisher, and the resulting measure of corpus size is much less accurate than the traditional word count.

Yet another solution might lie in normalizing the results against another linguistic feature. Because many news media databases do not provide much-needed linguistic tools such as counting the frequency of an individual item, part-of-speech tagging and punctuation sensitive searching, generating frequency data for a normalizing feature may very well turn into an independent study of its own. As an automated alternative, however, we can count the frequency of database entries (articles, editorials, press releases, broadcasts) containing tokens of certain features.

But which features should we normalize against? We might try using high frequency functional words such as the, in, and semi-functional words such as do (Table 2). However, not all database engines are alike, and some will cap search results at a certain number, as does Newspaper Source, effectively rendering useless counting entries with words like the.

<table>
<thead>
<tr>
<th>database</th>
<th>the</th>
<th>in</th>
<th>do</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newspaper Source</td>
<td>59,276,160</td>
<td>59,276,160</td>
<td>59,276,160</td>
</tr>
<tr>
<td>Canadian Newsstand</td>
<td>31,037,653</td>
<td>29,010,062</td>
<td>8,594,614</td>
</tr>
</tbody>
</table>

Table 2. Frequency of entries with functional words

©Universitat de Barcelona
Therefore, we should turn to somewhat less frequent words to detect variations in database size. General purpose verbs would serve that role well because they are still sufficiently frequent to generate meaningful comparisons; for example, the frequency of database entries that contain tokens of *give, take,* and *make* provide us with a meaningful comparison of database size (Table 3).

<table>
<thead>
<tr>
<th>database</th>
<th>give</th>
<th>take</th>
<th>make</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newspaper Source</td>
<td>2,004,365</td>
<td>4,197,623</td>
<td>4,367,807</td>
</tr>
<tr>
<td>Canadian Newsstand</td>
<td>3,196,769</td>
<td>6,219,011</td>
<td>6,922,226</td>
</tr>
</tbody>
</table>

Table 3. Frequency of entries with general purpose verbs

The comparability of the two databases can be seen in a very strong positive correlation \( r = 0.99 \) between the two sets of search results in Table 3. Further, when we calculate the normalization factors for the Canadian Newsstand, we get:

\[
\begin{align*}
give & \quad 3,196,769 / 2,004,365 = 1.6 \\
take & \quad 6,219,011 / 4,197,623 = 1.5 \\
make & \quad 6,922,226 / 4,367,807 = 1.6
\end{align*}
\]

These results are similar to the normalization factor of 1.5 calculated on the basis of days of coverage above. Given the cross-validation of the two normalization methods proposed here, we can be assured that our comparison of Newspaper Source and Canadian Newsstand will be statistically meaningful.

3.4 Procedure

I searched the mass media databases for tokens of \([be \{done, finished, started\}]\) NP. As a control condition I also searched for other tokens of the transitive *be* perfect with the participles *made, seen, heard, given, begun, ended, wanted, known, seen.* I used three criteria that helped me converge on this set of control verbs: 1) semantic similarity, 2) general purpose-ness, 3) prior documentation. First, I selected *made,*
begun, and ended because they semantically resemble done, finished and started; I avoided Romance verbs such as terminate and initiate, because formal Latinate vocabulary is unlikely to be retained in informal registers. Second, I selected given, made, seen, known because these are general purpose verbs, whose high frequency has been borne out by empirical studies in acquisition (2006). Considering their applicability to a large range of situations, one would expect that these verbs would be more likely to be retained during be perfect attrition. Third, I selected heard, wanted, and seen based on Wolfram’s (1996) and Gold’s (2007) attestation of these verbs in be perfect construction. I assumed, following Bybee (2006), that grammatical schemata are retained in memory along with lexical items; thus, there is a greater probability of these tokens resurfacing in the present-day transitive be perfect schema.

I entered the strings in (12) in the search engines. I excluded present-simple third-person singular forms because of potential homophony (he’s = he is or he has). As a final step, I manually identified tokens with direct objects NPs and excluded duplicate search results.

(15) {am, ’m, are, ’re was, were} {done, finished, started, made, seen, heard, given, begun, ended, wanted, known, seen}

The collected tokens were coded for participle choice, language user’s location, and year of occurrence. Information about language user’s location was intended to answer the question if the construction is restricted to any dialectal area in North America. Information about participle choice (done, finished, started) was intended to provide a clue about the productivity of the construction, as well as its subnational geographical distribution. Information about the year of occurrence was intended to clarify if the construction is an innovation or retention.

The returned results were checked for correlations with demographic data from official statistical sources, as well as another linguistic feature – the verb give. It was assumed that if [be done NP] is a general feature of either Canadian or American English, its pattern of subnational distribution would parallel that of give, a semantically general, dialectally unmarked verb.
4. Findings

The distribution of constructional tokens (Table 4) shows, with extreme statistical significance, that the construction is marginal in the US, and prevalent in Canada.

<table>
<thead>
<tr>
<th>database</th>
<th>country</th>
<th>population</th>
<th># tokens</th>
<th>normalized # tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newspaper Source</td>
<td>USA</td>
<td>313,900,000</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Canadian Newsstand</td>
<td>Canada</td>
<td>34,880,000</td>
<td>1217</td>
<td>1825.5^6</td>
</tr>
</tbody>
</table>

Population and tokens: N=2, r=-1, p<0.0001

Table 4. Distribution of tokens by database

The tokens have been geocoded and are available in interactive format both as a feature map at and as a heat map at http://yerastov.fhsu.edu/maps.html. Respective screenshots are provided in Maps 1 and 2. The maps demonstrate that the construction is spread across Canada, with hotspots concentrated in densely populated areas. However, raw frequency may be misleading if not weighted for factors such as population density or frequency of other linguistic features, a topic we turn to now.

Map 1. A heat map

^5 The population figures in this article come from the US Census Bureau and Statistics Canada at the time of writing.

^6 The normalization factor of 1.5 is used to adjust the Canadian result (see 3.3 for justification).
The results from Canadian Newsstand (Table 5 and Figure 1) show, with high statistical significance, that the construction is proportionately distributed across the Canadian provinces. When compared to the distribution of the verb give by province in Newsstand and the distribution of Canadian population, [be done NP] comes out as having an almost perfect positive correlation with those measures. Because of such close similarity in distribution with the general verb give (which one would presumably find in any dialect of English with comparable frequency), we can be assured that [be done NP] is a general feature of Canadian English. A positive correlation between give and provincial propulation validates the sampling method and confirms the internal consistency of data.
<table>
<thead>
<tr>
<th>province/territory</th>
<th>[be done NP] (Newsstand)</th>
<th>give (Newsstand)</th>
<th>population (Statistics Canada)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ON</td>
<td>531 44%</td>
<td>1,692,732 49%</td>
<td>13,505,001 51%</td>
</tr>
<tr>
<td>BC</td>
<td>292 24%</td>
<td>735,772 21%</td>
<td>4,622,001 18%</td>
</tr>
<tr>
<td>AB</td>
<td>150 12%</td>
<td>379,444 11%</td>
<td>3,873,001 15%</td>
</tr>
<tr>
<td>NB</td>
<td>104 9%</td>
<td>206,374 6%</td>
<td>756,000 3%</td>
</tr>
<tr>
<td>SK</td>
<td>66 5%</td>
<td>114,465 3%</td>
<td>1,080,000 4%</td>
</tr>
<tr>
<td>QC</td>
<td>35 3%</td>
<td>232,734 7%</td>
<td>1,059,001 4%</td>
</tr>
<tr>
<td>MB</td>
<td>28 2%</td>
<td>43,989 1%</td>
<td>1,267,000 5%</td>
</tr>
<tr>
<td>YT</td>
<td>9 1%</td>
<td>29,131 1%</td>
<td>36,000 0%</td>
</tr>
<tr>
<td>PE</td>
<td>1 0%</td>
<td>- 9 0%</td>
<td>146,000 1%</td>
</tr>
<tr>
<td>NT</td>
<td>1 0%</td>
<td>- 0%</td>
<td>43,000 0%</td>
</tr>
<tr>
<td>total</td>
<td>1217 100%</td>
<td>3,434,641 100%</td>
<td>26,387,003 100%</td>
</tr>
</tbody>
</table>

Table 5. Distribution of [be done NP] relative to give and population
(Canadian Newsstand and Statistics Canada)

---

7 The provincial codes are: ON = Ontario, BC = British Columbia, AB = Alberta, NB = New Brunswick, SK = Saskatchewan, QC = Quebec, MB = Manitoba, YT = Yukon, PE = Prince Edward Island, NT = Northwest Territories.
8 Anglophone population only.
9 Data not provided by ProQuest.
Noteworthy is the high number of constructional occurrences in New Brunswick, as compared to other provinces. There are more occurrences in New Brunswick than in Manitoba or Saskatchewan, even though it has a smaller population than either of the two provinces. Similarly noteworthy is the high number of constructional occurrences in two major New Brunswick municipalities – Moncton and Saint John (4% and 2% of the entire sample, respectively). If we take into account massive Scottish immigration to the Maritime Provinces in the 18th and 19th centuries (Bumstead 1981), then these high numbers can be attributed to a lasting founder effect. The absence of any tokens from Nova Scotia, another province with a strong Scottish influence, is simply due to the irregularity of representation of Nova Scotia sources in Canadian Newsstand (Proquest 2013).

A positive correlation can be found between the population size of a Canadian municipality and the frequency of [be done NP] (Figure 2). While somewhat weaker, this correlation is similar to the one between [be done NP] and provincial population, thus reinforcing the internal consistency of the findings.
[be done NP] and municipal population: N = 162, r = 0.72, p < 0.0001

Figure 2. Distribution of [be done NP] by municipal population

(Canadian Newsstand and Statistics Canada)

No instances of be perfect with the participles in the control condition have been found in Newspaper Source and Canadian Newsstand. The participles that are attested are finished, done, and started, e.g.:

(16) a. When the workers are done their programs, a lot of them stay and become undocumented.
   b. When you are done the game, you feel that you have gotten rid of a little stress.
   c. He said if drivers find themselves behind, drivers who are done their routes early assist other trucks.
   d. I am not exactly sure what this means when I first pick up the menu, and I'm not much further ahead when I am done my meal.

(17) a. When I am finished my training, I, too, will have no choice but to leave Quebec.
   b. By the time she was finished university she had been prescribed all major SSRIs [...].
   c. In Vancouver, when people are finished work, they are finished work.
   d. He's a very bright young man who was finished school and he was following his passion for skiing and the outdoor lifestyle.

(18) a. My dad passed away in 1995 when I was started my pro career in Finland and we ended up winning that year.
b. When I was started golf, Se Ri Pak won the U.S. Women’s Open tournament, so this tournament is really special for me.

c. You never want to put your team down in any situation but I was started that to stick up for a teammate.

d. Western Coal was started work in the Tumbler Ridge area five years ago in an effort to tap into a surge in world coal demand [...]

The participle *finished* dominates the Newsstand sample (64.5%), followed by *done* (35.17%) and *started* (0.25%). When assessing these findings, we should bear in mind that educated writers of English prefer to use lexically specific verbs to semantically general ones; regardless of dialect, many composition instructors have general discomfort with the form *done*. This prescriptive attitude appears to hold in Canadian English and might be skewing the distribution away from *done* towards *finished* in written media, including the sample under investigation.

The internal consistency of the findings can also be revealed when we examine the distribution of the participles by province (Table 6). Excluding *started* from consideration due to its statistical marginality, we can find a perfect positive correlation between the distribution of *done* and *finished* across the provinces (Figure 3).

<table>
<thead>
<tr>
<th>province/territory</th>
<th>done</th>
<th>finished</th>
<th>started</th>
</tr>
</thead>
<tbody>
<tr>
<td>ON</td>
<td>177</td>
<td>354</td>
<td>0</td>
</tr>
<tr>
<td>BC</td>
<td>86</td>
<td>202</td>
<td>4</td>
</tr>
<tr>
<td>AB</td>
<td>61</td>
<td>89</td>
<td>0</td>
</tr>
<tr>
<td>NB</td>
<td>47</td>
<td>57</td>
<td>0</td>
</tr>
<tr>
<td>SK</td>
<td>30</td>
<td>36</td>
<td>0</td>
</tr>
<tr>
<td>QC</td>
<td>8</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>MB</td>
<td>15</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>YT</td>
<td>4</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>PEI</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>NT</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>total</td>
<td>428</td>
<td>785</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 6. Distribution of [be done NP] by participle and province (Canadian Newsstand)
done and finished: N=10, r=0.98, p<0.0001

Figure 2. Distribution of [be done NP] by participle and province.
(Canadian Newsstand)

Canadian Newsstand provides attestations of the construction from 1978 up to the present moment. A rapid increase in occurrences around the year 1998 is most likely due to the proliferation of digital technologies, which made it easier to capture news media data.

5. Discussion of findings

In what follows, I will consider whether [be done NP] is a recent innovation in North American English and adduce additional evidence to support the claim that [be done NP] is marginal in the US and prevalent in Canada.

5.1 Retention or innovation?

The results of the present study show that the construction has been in Canadian English since at least 1978. Has the construction arisen through innovation in the last 40 years, or had it been present in North America before?
This question may be answered with the help of the Contemporary and Historical Corpora of American English. A search of these corpora returns 6 tokens of the construction:

<table>
<thead>
<tr>
<th>Token</th>
<th>Source</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our old Doctor Bates is getting a little out of date and he’ll be about ready to be put on the retired list by the time you are done your theological course.</td>
<td>The Witness (1917)</td>
<td>Grace Livingston Hill Lutz (1865–1947)</td>
</tr>
<tr>
<td>And it feels good to be finished the expedition because now I can breathe a sigh of relief that nothing really major happened to us during the expedition.</td>
<td>Interview with Pasquale Scaturro (2004)</td>
<td>Alex Chadwick</td>
</tr>
<tr>
<td>We found that Corny had not been mistaken about her influence over her family, for the next morning, before we were done breakfast, Mr. Chipperton came around to see us. He was full of Nassau, and had made up his mind to go with us on Tuesday.</td>
<td>A Jolly Fellowship (1901)</td>
<td>Stockton, Frank Richard</td>
</tr>
<tr>
<td>So you get into that bind. Coelho's response is, the Gore campaign will be finished its fund-raising.</td>
<td>CBS Face Nation (1999)</td>
<td>Jack Quinn &amp; Ceci Connolly</td>
</tr>
<tr>
<td>I am finished 2 chapters.</td>
<td>No Time for Sergeants (1955)</td>
<td>Mac Hyman</td>
</tr>
<tr>
<td>After Scarborough, I moved to Calgary with my family -- they moved there during the boom -- and I stayed there until I was done high school and then I went into women's studies in Concordia University in Montreal, believe it or not.</td>
<td>NPR interview with singer Lullaby Baxter (2000)</td>
<td></td>
</tr>
</tbody>
</table>

Table 7. Contemporary and Historical Corpora of American English: occurrences of [be done NP]

Additionally, a search of the database Literature Online and of the Gutenberg Project yields a further three instances of the construction:
"Why, Kalman," exclaimed his sister, "you are not half done your feast. There are such lots of nice things yet."

I never was brought up to wait on anybody but myself. I'll go down in the yard, and play with the big yaller dog, till they're done dinner. That's the curiosest dog I ever did see.---I can't find out whether his tail is cut off or driv in.

Eight! She must be done dinner by this time!

The results of these searches, combined with the results of the present study, lead one to construct the following timeline for \textit{be done} NP:

<table>
<thead>
<tr>
<th>Source</th>
<th>Author</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Green Mountain Boys</td>
<td>Joseph Stevens</td>
<td>1860</td>
</tr>
<tr>
<td>Frank Wylde</td>
<td>J. Brander Matthews</td>
<td>1880</td>
</tr>
<tr>
<td>A Jolly Fellowship</td>
<td>Frank Richard Stockton</td>
<td>1901</td>
</tr>
<tr>
<td>A Foreigner: A Tale of Saskatchewan</td>
<td>Ralph Connor</td>
<td>1909</td>
</tr>
<tr>
<td>The Witness</td>
<td>Grace Livingston Hill Lutz</td>
<td>1917</td>
</tr>
<tr>
<td>No Time for Sergeants</td>
<td>Mac Hyman</td>
<td>1955</td>
</tr>
<tr>
<td>Face Nation</td>
<td>CBS</td>
<td>1999</td>
</tr>
<tr>
<td>Canadian Newsstand</td>
<td>[multiple]</td>
<td>1978-2013</td>
</tr>
</tbody>
</table>

These occurrences suggest that the construction has been present in North American English for at least one and a half centuries. A retention account of \textit{be done} NP is consistent with external historical facts. The construction occurs in areas such as Canada, as well as Philadelphia and Vermont, where a Scottish founder effect has been
documented (Dollinger 2008; Bumstead 1981; Bennet 2003; Leyburn 1962; Shields 1996). Unsurprisingly, New England, Pennsylvania, New Jersey, and New York are the areas from which the United Empire Loyalists moved to found the country of Canada (Dollinger 2008: 64-76). It is likely that \[be \ done \ NP\] was a feature of the dialect substratum that the United Empire Loyalists brought to Canada, which may also have been reinforced by subsequent Scottish and Irish migration to the Maritimes (Dollinger 2008: 78-88) and to Vermont (Shields 1996).

5.2 Distribution in North America

The present study shows the marginality of the construction \[be \ done \ NP\] in the US and its prevalence in Canada. This finding accords with the (scarce) data from COCA, BCE, and the Strathy Corpus. A normalized comparison of these corpora is presented in Table 10. The comparison confirms that \[be \ done \ NP\] is predominantly a feature of Canadian English.

<table>
<thead>
<tr>
<th>corpus</th>
<th>country</th>
<th>corpus size (words)</th>
<th># tokens</th>
<th># tokens normalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>COCA</td>
<td>USA</td>
<td>450 mil.</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Strathy</td>
<td>Canada</td>
<td>60 mil.</td>
<td>7</td>
<td>52.5</td>
</tr>
<tr>
<td>CBE</td>
<td>Canada</td>
<td>1.5 mil.</td>
<td>3</td>
<td>900</td>
</tr>
</tbody>
</table>

Table 10. Corpus search results in synchronic corpora

The quantitative marginality of the construction in corpora of US English is consistent with its scarce attestation in the US. Fruehwald & Myler (2015) attest to the presence of \[be \ done \ NP\] in Philadelphia based on native speaker intuitions. Their attestation can be independently confirmed by voluminous anecdotes on the Internet. The construction has become the subject of multiple online postings by non-linguists who discuss the “correctness” of the construction in Pennsylvania. Thus, an anonymous Internet user notices a clash between the constructions \[be \ done \ NP\] and \[be \ done with
NP] in Pittsburgh vis-à-vis Philadelphia; in response, another anonymous user, who acknowledges the widespread occurrence of this construction in the Philadelphia area:

Quotation 1

[Question] How many of you use this grammatical construction -- "I'm done my homework." It is used by virtually everyone in the Philadelphia region [...] 

[Answer] [...] Yes, I come from the Philadelphia Region. If "I'm done my homework" or "I'm done the dishes" is said around here, it sounds completely normal. [...] 

http://www.antimoon.com/forum/t1301-0.htm

Another dialectal area for which there is limited documentation of the construction’s occurrence is Eastern New England. For example, two occurrences of [be done NP] were captured by the Digital Communities Project at Sterling College in the speech of residents native to Wolcott, Vermont (Orleans County):

(19) a. I’m all done school.

b. When he was done high school he was on his own.

Both northeastern Vermont and Philadephia have experienced a Scottish founder effect (Leyburn 1962; Shields 1996). For example, consider the case of Orleans County in Vermont, where several towns were settled by Scots (Shields 1996). After 1820, sixty Scottish families came from Scotland to settle in the towns of Crafsbury, Glover, and Greensboro. Pressured by the consolidation of agriculture in Scotland, they left their home, the upper Irvine Valley of eastern Ayrshire. Coming to Northeastern Vermont via New York and Montreal, they typically bought land from Yankees who were moving West (Shields 1996).

Linguistic reflections of such settlement history can be found in fictional literature. In the play The Green Mountain Boy (1860), the author uses [be done NP] in a series of Scottish linguistic features to build up the image of a local 19th century Vermonter:
(20) a. I'll go down in the yard, and play with the big yaller dog, till they're done dinner.

   b. A sample stands afore you.

   c. he has to git up about twelve o'clock every night Major, I understood you wanted to hire a chap; I s'pose a rale cute one.

   d. I can't find out whether his tail is cut off or driv in.

   e. I swow this grammar's awful hard stuff to larn.

The form afore ‘before’ occurs 1035 times in the Scottish Corpus of Texts and Speech; the form git ‘get’ – 371 times; the form rale ‘really’ – 29 times; the form larn ‘learn’ – 1 time. The form driv as a past participle has been attested as a feature of Scotch Irish by Montgomery (1997).

The geographical restriction of [be done NP] to select dialectal pockets with a Scottish founder effect suggests the likelihood of a Scottish etymology for the construction.

Further evidence supporting the Scottish etymology hypothesis may be found in the biographies of the authors (referenced in Table 9) who use [be done NP] in the speech of their fictional characters. These authors are, in one way or another, linked to locations with a documented Scottish founder effect: Canada, New England, Philadelphia, New York, and Georgia. Ralph Connor, born in Ontario to a family of Scottish ancestry, ministered in Alberta and Manitoba. Frank Stockton, born in Philadelphia, later in life moved to Nutley, New Jersey. Grace Lutz was a native of Wellsville, New York. Brander Matthews was literarily active in New York (although born in Louisiana); because the play Frank Wylde is set in New York, it is likely that [be done NP] reflects the language of that community. A native of Boston, Joseph Stevens Jones displayed a thorough knowledge of Scottish features in Vermont English in the play Green Mountain Boys. Finally, Mac Hyman was a native of Cordelle, Georgia.

The locations in which these authors were born or resided show consistent evidence of Scottish influence. Philadelphia has traditionally been a hub of Scottish migration (Leyburn 1962), where [be done NP] is attested today (see above); the occurrences in neighboring New York are thus unsurprising. Georgia was colonized by
Scots as early as in 1736 (Parker 1997). As discussed above, Northeastern Vermont experienced a similar Scottish founder effect – at least in the early 19th century (Shields 1996). And Canada has been influenced both by the substratum of Scottish features of the United Empire Loyalists, as well as later 19th century Scottish immigrants (Dollinger 2007).

The restricted geographical distribution of \([ \text{be done } \text{NP}] \) goes hand in hand with its marginal sociolinguistic status. While there is widespread acceptance of the construction by Canadian editors, the construction seems to have been stigmatized as nonstandard in the US. Particularly illustrative is a prescriptive assessment of the construction by Capital Community College in Hartford, CT, in its Guide to Grammar and Writing:

Quotation 2

"I am done my work" is completely unacceptable; I can't imagine any level of discourse where it would be tolerated. "I am done with my work" is surely acceptable in an informal setting; "I have finished my work" would be an improvement, of course.

http://grammar.ccc.commnet.edu/
grammar/grammarlogs2/grammarlogs306.htm

Such editorial taboo likely leads to \([ \text{be done } \text{NP}] \) being edited out of print US media – hence its non-occurrence in \textit{Newspaper Source}.

6. Conclusion

I have shown on the basis of the present study that the construction \([ \text{be done } \text{NP}] \) is a century and a half old North American retention, prevalent in Canada and marginal in the US, where it is restricted to a few dialectal pockets. I have also shown that in fictional literature the construction co-occurs with other sociolinguistically marked Scottish features and that fiction writers’ biographies have in common links to areas of documented Scottish influence. I have argued, on the basis of attested dialectal parallels
that the construction [be done NP] has a Scottish etymology, and is probably a reflex the transitive be perfect in insular Scots.

The results of this study confirm the generalizability of the experimental findings based on the Illinois and Alberta samples (Yerastov 2012, 2010) to Canada and the United States. While both experimental and corpus studies have their inherent methodological drawbacks, the convergence of their results contributes to our confidence that we have adequately described the distribution and properties of the construction.

References


