‘she bes delighted with herself’ – Habitual marking in Irish English

Av: Hugh Curtis
Handledare: Harriet Sharp
### Table of Contents

1.0 INTRODUCTION AND AIMS..............................................................................................................3

2.0 BACKGROUND..................................................................................................................................4
   2.1 Aspect and Tense............................................................................................................................4
   2.2 Difference between continuous and progressive aspects and the simple form ..................8
   2.3 Habitual aspect in Standard English............................................................................................9
   2.4 Habitual aspect in Irish English....................................................................................................11
   2.5 Historical background of the habitual markers in IrE.................................................................12

3.0 MATERIAL AND METHODS...........................................................................................................17
   3.1 The Electronic World Atlas for Varieties of English...............................................................18
   3.2 WebCorp Linguist's Search Engine.............................................................................................18
   3.4 A Corpus of Global Web-based English, GloWbE.................................................................20
   3.5 Advanced web-based search or crawl of .ie social media.........................................................20

4.0 RESULTS.........................................................................................................................................22
   4.1 eWAVE The Electronic World Atlas for Varieties of English: data and results..............22
   4.2 Results from the WebCorp Linguist's Search Engine data.....................................................25
   4.3 Results from the GloWbE data....................................................................................................27
   4.4 Results from advanced web-based searches of .ie social media...........................................29

5.0 CONCLUSION.................................................................................................................................32

REFERENCES...........................................................................................................................................35
ABSTRACT

The habitual aspect has been a feature of Irish English for centuries. How it has evolved may have had a lot to do with contact between Standard English and the Celtic language, Irish, spoken in Ireland. As time passes does the impact which these two languages have had on each other weaken? How has a major feature of Irish English, the habitual aspect, fared in the digital world? This essay executes some digital detective work and finds that habitual markers do be always there…

KEYWORDS: Irish-English, habitual aspect, corpus studies, Twitter, Irish English.
1.0 INTRODUCTION AND AIMS

Vernacular Englishes, a term used in sociolinguistics to refer to the indigenous languages or dialects of speech communities throughout the English speaking world (Crystal 2008: 511), may be in decline given the access speakers now have to Radio, TV and the Internet where standardised English dominates. The variety of English spoken and written in Ireland (Latin: Hibernia) has been referred to variously as Hiberno English (HE), Anglo-English and simply Irish-English (IrE). Since the Middle Ages, Hiberno English has been spoken in Ireland where it is understood to have survived in an unbroken form at least within the Pale (Hickey 2000, Bliss 1979), the historical area on the East coast of Ireland, stretching from Dalkey, just south of Dublin, to Dundalk on the border of present day Northern Ireland. Although different names exist for this language variety, it will here be referred to as Irish English (IrE).

IrE is the oldest overseas form of English. As well as being influenced by settlers from England and Scotland, the native Gaelic, a Celtic language, has influenced Irish English, more so in some periods than in others. The 19th century for instance saw a drastic demise in Irish: its abandonment by the native Irish in favour of English. This concerted movement from Irish to English as the language of everyday speech occurred over one short generation and it impacted how English evolved in Ireland. (de Fréine 1970:83) In earlier periods English had made no headway at all and was even threatened with being engulfed by the Irish language on all levels. Although the official language of the Republic of Ireland is Gaelic, or ‘Irish’ as it is called in Ireland¹, English is today the everyday language of 98 percent of the population.

This study focuses on particular ways of expressing habitual aspect in IrE. A wider range of asceptual distinctions are realised in Irish English than in Standard English (StE). To understand these distinctions the relationship between aspect and tense must be examined and this will be done in chapter 2. IrE marks aspects in several ways. For instance, to mark habitual or extended action IrE uses the verb ‘be’. Sometimes a non-standard ‘s’ is added forming ‘bes’, or ‘bees’, or ‘be’s’. Habitual or extended action can also be marked with ‘do’ on its own or together with ‘be’ forming ‘do be’. Not be confused with emphatic or contrastive ‘do’, which are found in other Englishes, this habitual form has ‘do’ used periphrastically. Examples of ways of realising habitual aspect in IrE are shown below.

¹ Romaine in the introductory chapter of Nic Pháidín and Ó Cearnaigh eds (2008:6) estimates that the active Irish-language scene probably comprises only 5 to 10 per cent of the island’s population, and around one in three people (c 1.8 million) on the island can understand Irish to some extent.
The aim of this essay is to attempt to find answers to the following qualitative and quantitative research questions:

a. How are habitual aspect markers used in IrE?

b. In what linguistic contexts is the habitual aspect realised primarily?

c. How common are habitual aspect markers in a specific corpus or internet forum?

d. What type of speakers use these forms primarily?

The data examined in order to answer these questions come from a number of corpora listed in chapter 3 (Material and Methods).

It is hypothesized that the feature of habitual marking with ‘be(es)’ and ‘do(es) be’ in Irish English has not receded but, on the contrary, remains a vital element of IrE at the interface where people use the language in their daily lives. The feature will have very low frequencies however in standardised Irish English.

2.0 BACKGROUND

Given the linguistic nature of this study and its focus on a particular aspect of a vernacular form of English that may not be completely familiar to the reader, it is necessary to contextualise the matter. Instances of the phenomenon will be brought up against a backdrop of not only the forms of grammar having to do with the subject but also the background history of IrE and the previous studies conducted on habitual aspect in IrE.

2.1 Aspect and Tense

Tense and aspect are often confused. Whereas tense is the temporality or ‘time’ when a verb is ‘performed’, aspect shows the temporal structure of the ‘activity’, ‘event’, or ‘state’ (Richards et al 1992: 34), and whether it is ongoing or completed. Aspect has primarily to do with how the grammar marks the length or type of temporal activity that the verb indicates

---

2 Entry in the ICE-Ireland corpus (<ICE-NIFTFS1A-032$A><#>Hejuststandsthereand){}<>
Aspect will typically carry information about habituality, duration, repetition, punctuation, etc. Aspect is not so much a category of grammar but is a term used to distinguish between general semantic oppositions possible (Comrie 1976: 6).

Spatial relationships is how people conceive time, and concepts dealing with the notion of language’s relation to the world of events and actions are thought of as being relevant to the speaker in terms of space. A common notion of time is that it can be sometimes compressed, sometimes drawn out. In a crude sense English has only two tenses; a non-past ‘present’, mostly used to talk about the future and the now, and a past tense (Trask 2007:295). All other characteristics are expressed through aspect. The future is not marked by verbal inflection in English but by modal verbs such as will and shall. Tense, the grammatical category that relates to time, is nothing but a grammaticalisation of time (Trask 294). With no future tense English expresses future events in relational non-past ‘present’ tense through the medium of what is termed aspect. Aspect is an internal temporal structure independent of an event’s actual position in time.

Being prone to see the world in the framework of a ‘timeline’ (Michaelis 2006: 220) speakers refer to being able to ‘head towards’ the future or ‘returning’ to the past. This timeline of past, present and future, or a section of points in time, is a common concept for most speakers of English (other languages have more tenses). Since English uses aspect to express various characteristics of what can be expressed, the relation of tense to aspect in English is complex, but these relations are not inextricable at the semantic level. Categories can be unwound in order to differentiate tense from aspect (Michaelis 2006: 241), these relations are not inseparable (Comrie 1976: 6-7). One way to approach this difference is through the concepts of code time, the moment of utterance or composition, reference time the explicit point in time being referred to, and event time, the status or duration of the event. The analysis of a phrase can thus take in the code time or point of speech (the now), the point of reference (the perspective from which the event is considered), and the point of event (the precise moment when the event took place) (Reichenbach 1947: 287.298). We mentioned aspect reflecting the internal temporal structure of the event, tense points out the time of the event as a whole as it relates to that reference time. Reichenbach’s ‘reference time’ is “the time for which, on some occasion, a claim is made (Klein 1992: 535). Aspect provides information about the internal structure of the event itself, whereas tense is deictic— locating the point in time relevant to the point of time of utterance (Comrie 1976: 2).

By way of introduction to habitual aspect marking in IrE, I will be using notions of aspect in this essay set down by Bybee, Perkins and Pagliuca in The Evolution of Grammar.
Tense, Aspect, and Modality in the Languages of the World. (1994:125-175). Their work sets aside former structural notions of opposites in the realm of aspect for instance. Using the idea of ‘Grams’ grammatical morphemes (bound or free), as a base unit for understanding and analyzing features, especially the traditional category of aspect, and tracing the evolution of universal features in a theory of Language structure. I will not be using however the term ‘anterior’ which is the suggested Bybee et al term for what is commonly referred to as perfect. The justification for using this term is to avoid confusion between ‘perfect’ and ‘perfective’, distinctions which are more commonly understood in the Slavic languages for instance where distinction is marked in morphology between perfective/imperfective. There are verbal forms of the constitutive parts of the verb phrase that mark an ongoing internal temporal sense. Within this realm of temporality there are either verb phrases which mark the temporal aspect of the imperfect in an overt manner, or phrases that do not readily do so. Verb phrases used to mark progressive on-going actions usually utilize dynamic verbs (such as run, drink, walk) whereas static verbs, such as ‘be’ for instance, are used to stress the habitual element. These are commonly used in Ire and are grammatic(al)ized in the language. These latter aspects are called habitual markers, they denote an ongoing habitual sense and are to be found in Irish English and are commonly referred to as ‘habitual be’.

Aspect is often ambiguous in English, the sentence ‘Paddy is reading Ulysses’ can mean that the subject has the book in his hand at the moment, or that he is underway with the activity i.e. that he is not necessarily holding Joyce’s work in it in his hand but he is in the process of reading it as such. Frequency adverbials, such as usually, always, normally, etc., are the common markers of habitual aspect in StE. Otherwise present tense is used in an unmarked manner.

(5) He is usually behind bars

Perfective and imperfective aspects are usually the defining categories in English. The former is the aspect of a bounded situation, the homogeneous non-temporality of which is delimited. The imperfective on the other hand relates to temporal situations existing continuously or repetitively. There are additional distinctions such as states and ongoing actions separating continuous and progressive aspects from repetitive or habitual aspects. This so-called progressive aspect denotes ‘a singular present event’ (Crystal, 2006:147), and can be seen to have a relationship to the habitual aspect since it can be argued that it ‘overlaps’ the domain of the habitual aspect to some extent. Both aspects are imperfective, in the sense that their
operational area is not completed or temporally bound. The operative domain of the so-called progressive aspect begins in the sentence below at the point where the ‘been’ emerges.

(6) The student might have been writing a letter.

Two types of aspect are usually recognised in English. The first type is Aktionsart or **lexical aspect** (or inherent lexical aspect). In this type, the semantic meaning is internal to the verb itself. This is illustrated in the distinction between the verbs ‘nibble’, ‘eat’, and ‘devour’ (Trask 2007:27). Lexical aspect can be divided into four sub-categories with its own type of verb. This is described in Table 1 below.

<table>
<thead>
<tr>
<th>Sub-category</th>
<th>Type of verb</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. States</td>
<td>stative verbs</td>
<td><em>be, possess, have, want, lack</em></td>
</tr>
<tr>
<td>2. Activities</td>
<td>dynamic verbs</td>
<td><em>run, drink, walk</em></td>
</tr>
<tr>
<td>3. Accomplishments</td>
<td>durative verbs</td>
<td><em>run a marathon, read a book</em></td>
</tr>
<tr>
<td>4. Achievements</td>
<td>non-durative verbs</td>
<td><em>finish, accomplish, arrive, complete</em></td>
</tr>
</tbody>
</table>

Verbs in the first category, ‘stative verbs’, express unchanging conditions such as ‘to be’ or ‘to possess something’. It can be argued that these verbs stress the habitual element of a state that continues over a period of time. The second category contains ‘dynamic verbs’ which mark progressive on-going actions such as ‘run’ and ‘drink’, which have no inherent beginning or end. Verbs of the third category, ‘accomplishments’, are durative, i.e. they (last for a period of time and they have an inherent beginning and end e.g. ‘run a marathon’ or ‘read a book’. Verbs of the fourth category, ‘achievements’, are non-durative, i.e. they have a specific index in time, e.g. finish and arrive.

The second type of aspect is **grammatical aspect**. This type uses the grammatical tools at the disposal of a language, such as verbal auxiliaries, prefixes and suffixes, to encode the various outlooks that a speaker wishes to have towards actions, events and states. Grammatical aspect in English has two realisations (morphologically): (1) the progressive aspect and (2) the perfect aspect. These aspects work together with tense to yield various temporal sequences.

Grammatical **tense** is the use of inflections or auxiliary verbs to show that some state or action occurred along a timeline of past-present-future, i.e.; before, simultaneous with, or after some reference point, as shown in example (7):

(7) I ‘ate’ - I ‘eat’ - I ‘will eat’.
Grammatical aspect, in contrast, refers to the time when the action of the verb occurs either across or between those periods; it is the temporality of the event i.e. the times on the time line can be extended (progressive aspect) or combined (perfect aspect). In the progressive aspect, the action is continuous:

(8) When James called, I was eating.

In the perfect aspect, the action happened at some unknown time between then, the past, and now, the present time of speaking:

(9) I have eaten today.

Tense describes whether an event takes place in the past, present or future. Aspect describes the duration that the event covers; the event’s intrinsic temporality.

2.2 Difference between continuous and progressive aspects and the simple form

The categorisations of tense and aspect have been discussed above and given the limitations of this paper however only some grammatical categories will be touched upon in order to introduce habitual markers in IrE. Firstly a discussion of the difference between the continuous and progressive aspects and the simple form.

In StE the continuous and progressive aspects are deemed to be one and the same and are many times commonly referred to as the continuous aspect, although Comrie (1976:12) insists that the latter is a subdivision of the former. According to Quirk et al (1985:197-198), the progressive can be separated into three components:

1. the happening has duration
2. the happening has limited duration
3. the happening is not necessarily complete

Typically formed with an auxiliary verb, such as be, + ing: the present participle, the progressive shows that an action or state, past, present, or future, was, is, or will be unfinished at the time referred to, as shown in examples (10-11) below:

(10) I am listening to the Rolling Stones.

(Action still incomplete at this point in time)

(11) He was having his lunch when the dog barked.
(Action still incomplete at the time the dog barked).

By contrast the simple form, the bare verb stem with no inflection (also called the base form) is used for the infinitive in English. It functions as the simple present form as well as the imperative.

(12) I watch Tom’s car.

(13) Watch my car!

This base form is the form of the verb found in dictionary entries, and it functions as the subjunctive mood for most persons including the third-person singular.

2.3 Habitual aspect in Standard English

We have mentioned how the habitual aspect is not usually overtly marked in StE. There is an inherent generic meaning in a habitual present such as in ‘Water boils at 100 degrees Celsius.’ Such an action/state is not only generic or gnomic but it holds true for all times (Bybee et al, 141, 152). Modal ‘will’ also conveys this generic sense. The sentence above is continuous even though the water has passed its boiling point the statement still continues to be true/known. Habitual present is a term applied to a morphologically unmarked present which speaks two events occurring on a regular basis:

(14) She stays up till about half past five gets up at nine every day.

(Meyer 2009:189)

This inherent habitual aspect which is part of the semantic makeup of the present tense, morphologically unmarked, is one of the reasons why some scholars, (Quirk et al, 2007 & Trask 1985:295) included, prefer to used the term ‘non-past’ to denote the present tense in English. The sentence ‘the man works every day’ indicates more aspect then tense, given that it does not mark a point in time in the deictic sense (as is the case with tense) but a habitual action (Meyer 2009:189). This ‘hidden’ aspect in the English makes it difficult to exclusively determine present habitual markers in any data available in StE. The habitual aspect in StE is demonstrated in an unmarked form in English in the following sentences which are commonly referred to as past, present, and future (Comrie 1976: 25, 30, 98-99, 114-115, 124):

(15) I walked to work every day for ten years

(16) I walk to work every day

(17) I will walk to work every day after I get well
In the past tense a separate habitual aspect is marked by the phrase ‘used to’ or the ‘used to construction’ such as ‘John used to work here.’ Comrie (1976: 25). But the seemingly ‘overlapping’ progressive ‘John was working (when I entered)’ suggests that a state is continuous in an internal temporal sense. These are arguably separate forms of the habitual in English but they only exist in the past tense, otherwise, as Comrie notes, there is just the simple form ‘John worked here’ with no particular marking to denote the habitual sense, and only the progressive sense is excluded. Binnick argues that the periphrastic form ‘used to’ when used in stative expressions (16) does not receive a habitual interpretation (Binnick 2006:254). This statement is in direct contradiction to Comrie on the matter (1976:28).

The simple and progressive forms are more straightforwardly separable in other languages such as the Romance languages. Comrie states that the habitual and progressive aspects too can be combined in English, such as in the sentence ‘He used to be playing’. The habitual describes a characteristic feature of an extended period of time (28), whether or not this period is iterative or not (i.e. broken up into several spots or instances). Habitual aspect can combine readily with the progressive aspect in StE as in ‘when I’ll visit John he’ll be playing the flute’.

Apart from the periphrastic ‘used to’ the auxiliary ‘will/would’ can also mark habituality in English such as in the sentences below.

(18) Every now and then, John will/would suddenly burst into song. (Michaelis 2006: 248)

(19) He will make mistakes won’t he?

The habitual aspect is not overtly marked in StE in the present tense. Particular past tense habitual markers exist in the ‘used to’ and ‘will/would” constructions. But some contention exists as to the bounds of the habitual aspect in StE. Arguments by Binnick (2005) refute all traditional habitual constructions with ‘used to’ arguing that it is neither a state nor a cessation of a state, nor is it even a past tense but a kind of present perfect. Even the means of the simple present or past tense cannot carry the meaning of habitual aspect according to Binnick, they can only receive habitual meanings in a contextual situation (Binnick 2005:340). Only habitual marker in English is the modal ‘will’ and would being, according to Binnick, simply its past tense.
2.4 Habitual aspect in Irish English

The marking of habitual aspect is much more strongly grammaticalized in many varieties than it is in Standard British or American English. The most widespread habitual markers are be (invariant as in African American Vernacular English He be sick, or inflected as in Irish English He be's at home), does/doz (in practically all pidgins and creoles, e.g. Barbados (He does catch fish pretty), or combinations of the two (e.g. Irish English There does be a meeting of the company every Tuesday). Be or do be are often also combined with the Progressive in marking habituality (e.g. African American Vernacular English I always be playing ball or Irish English They do be shooting there a couple of times a week). (Kortmann 2006:607)

The habitual aspect in IrE denotes an ongoing habitual relation to and action or state. Unlike StE habitual aspect is overtly marked morphologically in the present tense in IrE. One IrE feature of the habitual aspect is commonly referred to in the literature as habitual be. Often the habitual be takes an inflectional ending –s, forming ‘bes’, as can be seen in example (20) below.

(20) She bes on Facebook on her blackberry

The meaning that the inflected form of ‘be’ carries here is that the subject is usually on Facebook via her Blackberry phone (periodically or habitually). This example is non-standard English. The example is taken from a Twitter user, and it is possible to find out much information about the ‘speaker’. In this case, the speaker is a young urban woman of 18 living in her hometown of Derry, in Northern Ireland. This habitual be form is a common form in IrE and has been verified in the literature by scholars such as Bliss (1976, 1977, 1979), Corrigan (1993), Hickey (1993 1997, 2007), Kallen (1997, 1994) and (Odlin 1991,).

Another use of ‘be’ which is seen recurrently in IrE, and one that has a similar sense to Habitual be is the ‘does be/do be’ construction that is used to indicate the continuous, or habitual, present. No consensus exists in the literature about how this construction should be named, but for this study it will be referred to as the habitual does be form. This construction is exemplified in (13):

(21) She does be on Facebook on her blackberry

Scholars working in the area (Kallen (1994, 1997), Hickey (2007), Corrigan (2010) Ronan (2011) and Filppula (2008) stress that much remains ‘unsettled’ as concerns the categories of this phenomenon of the habitual be. Its history and genesis are also disputed, but Ronan
(2011: 105-118) states that there is general agreement that there have been two converging directions of input which have led to the emergence of the phenomena in IrE. Citing Ihalainen (1991) and Klemola (2002), Ronan argues that some traditional British English dialects also have a phenomenon called **periphrastic do** as a habitual marker, where it seems to have been retained; basically having survived from south western dialects of English. The other potential contributor to the origin of present habitual aspect marking in IrE is contact between earlier English and Irish Gaelic. Hickey and Kallen are in agreement with Filppula et al. (2008: 190) on this point). Filppula sees the contact with Irish as the source of both the do(es) be and the be forms. Section 2.4 will discuss the historical origin in greater detail. The term ‘periphrastic do’ simply refers to constructions with the auxiliary verb *do*. Since verbs in English are not heavily inflected, combinations of tense, aspect and mood are realised periphrastically using constructions with auxiliary verbs and modals.

A major factor in how IrE speakers have determined their unique linguistic identity in the past is by using this periphrastic do-construction. The term periphrastic is descriptive of one or more free morphemes carrying meaning in a sentence, as opposed to the verb receiving inflected affixes or through derivation). Writers, such as Synge and Yeats, and even Bram Stocker have used this construction in many of their works. This study will attempt to find out if habitual aspect markers are still used in contemporary IrE.

### 2.5 Historical background of the habitual markers in IrE

Although the combinations of linguistic elements forming the habitual aspect in Irish English are few, a great deal has been written by scholars interested in the subject. This section will give an overview of previous scholarly work on the habitual aspect in IrE, and indeed the earlier terms for the language itself. The habitual aspect in IrE has been debated and commented on as far back as O’Neill (1947) and Henry (1957) when Irish English was commonly referred to as **Anglo Irish**, now an abandoned name. With reference to earlier studies Kallen (2012: 31) stresses that much of the previous debates concerned this unsettled question about how to differentiate between the various subsets of the ‘generic/habitual verb forms’. In particular, he sees the problem as one of how to differentiate within the sets, or subdivide the various forms i.e. *do(es) be, bes, bees, be’s* and *do*. Shimida (2013:1) writes that the problem of locating or mapping Irish English forms has become extremely difficult in recent times due to speaker awareness of “bad grammar”. Also, modern media function as a
homogenising factor which might mix dialects which were formerly isolated, extending the
frontiers of standard English. (Chapter 3 of this study will also touch on this issue.)

It is clear that scholars disagree about the genesis of forms of the habitual aspect in IrE.
The situation is made worse by the fact that there is no clear division of territory for the
different forms. Kallen (1994:180) notes that the ‘be’ forms are more prolific in the Ulster
dialect than in the south. This dialect is spoken in the of Ireland and it is associated with the
influx of primarily Scots speaking peoples, and the historical plantations of the 1600s which
displaced many members of the native Gaelic speaking population and the native aristocracy.
(The Northern dominance of the be(es) form has also been observed by Filppula 2004: 79,
maintains that this Ulster Scotch element is a possible location for the genesis of the ‘bes’
form (2007), but he adds the fact that many instances have been noted in Co. Wexford in the
southeast where an archaic form of English, so-called Anglo-Normal English, Yola3 or the
dialect of Forth and Bargy thought to have evolved from Middle English and to have survived
up to the late 17th century.

Hickey (2007) comments on the fact that attestations of ‘be(es)’ are few and far between in
Southern Irish English (SIE). One of the first mentions of the use of the habitual aspect occurs
in 1845 in O’Donovan’s A Grammar of the Irish Language. O’Donovan writes that the Irish
attempt to produce in English an equivalent to the Irish habitual tense which he refers to as
the Consuetudinal Present. ‘The Irish attempt to introduce this tense even into English, as “he
bees”, “he does be”, &c.’ (O’Donovan 1845: 151). Hickey sees as widely significant the fact
that a mid 19th century grammar, albeit an Irish one, mentions ‘bees’ explicitly and in the
sense that the Irish natives are attempting to rope in, a term that will work for the purposes of
new learners of English who might otherwise feel that English lacked this tense Hickey

Hickey (2007:227) lists a great many possible theories for the prolific use of bes in Northern
Ireland. He points out that various scholars prefer a scenario where Irish speaking natives
adapted the ‘bees’ term from Ulster Scots use, where it had been a relic from a Germanic
differentiation between the generic ‘wesan’ and habitual ‘beon’ (McCafferty 2007, qtd. in
Hickey 2007: 227). This was later replaced by ‘is’ on all forms of Scots origin (Traugott
1972: 89, 191f.) Hickey (2007:227) also finds it a remarkable ‘fact’ “…that neither structure,

3 Meaning ‘old’.
4 Also quoted in Montgomery and Kirk 1996.
do(es) be nor be(s), is significantly attested before the Nineteenth Century, indeed the latter is not documented at all before the middle of that century.”. Filppulla (1999:144 ff.) and Montgomery (2006a: 82-3) are in agreement with Hickey too that the forms ‘do(es) be’ and ‘be(es)’ show an element of influence from Irish, in which punctual and habitual forms of the verb are important contrastive categories. See examples (22-23) below.

(22) Tú si
    BE+punctual she = ‘She is’

(23) Bionn si
    BE+habitual she = ‘She usually is’

Filppula et al. (2008:190) acknowledges the substratal genesis for both the ‘do(es) be’ and ‘be(es)’ forms of habitual in IrE. The terms substrata (or singular substratum), and by extension superstratum, stem from applied theory of language contact. One language coming into contact with another has a lesser status and gives way in the culture to the new more socially prestigious or economically sustainable language loaded with prestige or status. The undercurrent, or substrata, of the language abandoned still exerts influence from ‘below’ on the newly acquired language. Superstrata, in contrast, exert influence from ‘above’. In the case in question here, Filppula et al (2008:190) argue that in the cross-over from Irish to English, the form of the ‘Consuetudinal Present’ of the early Modern Irish (the Irish grammarians’ term for the habitual) was the substantive verb ‘be’, the 3rd person singular forms of which were ‘bídh’ and ‘bí’ respectively. Filppula et al also add that the existence of the special form in the habitual aspect in the Irish language that is audibly similar to ‘be’ makes a claim for substrata influence clear enough.

Regarding tense and aspect (which was discussed in section 2.1) Hickey (2007: 213) propounds that there are several ways in IrE for aspect, in verbal expression, to go beyond tense to convey some ‘extra information’. Hickey divides the present habitual aspect into durative and iterative. The durative form “characterises a repeated action which typically lasts for a certain length of time” (Dahl 1985: 95-102 qtd. in Hickey 2007). The iterative form, in contrast, is used when the durative nature is less important and stress is placed on the repetitive nature of the action. In StE the habitual aspect can be expressed by means of the present tense forms — simple present for iterative uses and continuous present for durative uses — that are often accompanied by adverbs such as always, often, frequently, etc. (Hickey,
In addition to standard forms, IrE can express the present habitual aspect with a special category of the verb phrase without having to resort to doing so periphrastically with always, often, etc. According to Hickey (2005: 28, 2007: 214-232) these are:

a) the “suffixal –s on lexical verb stems” expressing the iterative habitual;
   *I gets all mixed up with the buttons on the recorder* (Hickey 2007: 214-215).
   (Occurring in southern Ireland)

b) do (es) be expressing the durative habitual;
   *She does be reading books* (Hickey, 2005: 28).
   (Occurring in southern Ireland)

c) be (es) expressing the durative habitual;
   *They bees up late at night* (Hickey, 2005: 28).
   (Occurring in Northern Ireland)

In terms of superstrata coming down to match this substrata rising from the Native Irish below, Hickey sees, in a text by John Michelburne called “*Ireland Preserved, or The Siege of Londonderry*” (1705), a South Western English influenced form originating among settlers. Two different forms express the habitual in this text:

1. ‘do’ + lexical verb, which also occurs in South-West England
2. ‘do be’ + infinitive form

Claims could be made, according to Hickey, for the second form being an “intermediary” between the habitual ‘do’ + lexical verb and the present-day Irish English form of the habitual ‘does be’ + V-ing:

```
+ V-ing          + bionn + non-finite verb form
```

(24) *They do be fighting a lot.  Bíonn siad ag troid go minic.*
    (Hickey, 2007:134) [is- they at fighting often]

Hickey (2007: 214-215) gives the following example of a suffixal –s on lexical verb stems expressing the iterative habitual and occurring in southern Ireland:

(25) *I gets all mixed up with the buttons on the recorder.*

---

5 Hickey also comparisons the habitual aspect in new world Englishes with the Irish vernacular which also has a formally encoded habitual. He proposes a link between African American English by way of possible close interaction between Irish natives and peoples of the Caribbean speaking various types of English creoles.
The implication of instances such as (1) above is seen by Hickey (2007: 220) to be that ‘do’ + lexical verb had habitual uses in Early Modern Irish English. If this interpretation is correct, then the source of this ‘do’ + lexical verb habitual would have been the result of South Western British English input to the east of Ireland which subsequently spread to the rest of the country. Corrigan (1997b) sees this input too from Irish substrate, coupled with forms from English and Scots, to be the origin of the ‘be(s)’ habitual in east Ulster in the mid 19th century.

Hickey (2007: 214-215) states of his data that “[s]peakers would appear to exploit the contrast of an iterative and a durative habitual.” He goes on to illustrate a fine example of how the durative habitual can contrast with the iterative. He offers the following examples. In (25) we see an instance of the iterative:

(25) I looks after the little one for her then (2007:215)

If this sentence has been in the habitual, it might have been put:

(26) And I do be looking after the little one

A parallel sentence stressing duration is shown in (18):

(27) And I do be dying for sweets (2007:215)

Although Hickey sketches out an elaborate schemata of subcategories under the habitual aspect, other scholars have not readily taken up his terms in their own analyses. This problem is compounded by the fact that there is no clarity concerning the geographic spread of the various forms, not to mention their origin. However, all scholars agree that there are two directions of input to the habitual aspect in IrE: (1) that of Irish substrata and (2) English retention of some past form, like periphrastic ‘do’, which had grown redundant or semantically empty in the language. This form was then appropriated by the native Irish as a way to mark grammatical structures which they found necessary in the new language, structures that existed in the old language, Irish, which they were rapidly rejecting in the mid 19th century.

This brings us to the present century where a great deal has been written on this subject. Yet in contrast, in an analysis of the ICE-Ireland Corpus, Ronan (2011: 111) finds extremely low instance(s) of the ‘do’ form of the habitual aspect marker. In fact, in a million words, she finds only one instance and only three ‘be(es)’. She does state, however, that the few instances of all vernacular features in the corpus as a whole leads her to believe that there
are many inhibiting factors at play here, not least of which is the fact that the larger part of the corpus is made up of written standardized, non-dialectal language use. It is no doubt the case that more instances would turn up in a corpora homing in on vernacular speech. What follows is a presentation of the material and methods used in the current study.

3.0 MATERIAL AND METHODS

This chapter describes the material and methods used in researching this paper. The method used was a corpus investigation delimiting the phenomenon as well as evaluating the difficulties met with when undertaking this kind of corpus evaluation of a single feature—habitual markers. Both quantitative and qualitative methods were used in accessing the data and proceeding with the analysis. Much difficulty was encountered when attempting to gain access to certain corpora. All attempts to gain access to any spoken corpus of Irish English were fruitless. What follows is a short list of all materials, tools and methods used to gain access to and extract the data. Specific materials used were the following:

I. An advanced web-based search or crawl of social media and popular forums in the .ie country-code top level domain (Ireland). The latter concentrated on advanced searches of semipublic language available on the social networking service Facebook and the so-called mini-blog Twitter.

II. A Corpus of Global Web-based English, GloWbE, composed of 1.9 billion words from 1.8 million web pages in 20 different English-speaking countries. Data was collected from the (entire island of) Ireland section which has one million words. Created by Mark Davies of Brigham Young University, the corpus was released in April 2013.

III. The on-line WebCorp Linguist’s Search Engine out of Birmingham University with access to the following corpora.

A. Synchronic English Web Corpus: 470 million word corpus built from web-extracted texts.

B. Diachronic English Web Corpus: 130 million word corpus randomly selected from a larger collection.

C. Birmingham Blog Corpus: 630 million word corpus built from blogging websites.

IV. eWAVE: the electronic World Atlas of Varieties of English, a web based international interactive data language base run by the Freiburg Institute for
Advanced Studies (FRIAS) and the English Department of the University of Freiburg, Germany.

3.1 The Electronic World Atlas for Varieties of English

In order to achieve an international backdrop for presenting the habitual feature ‘be(s)’ and ‘do(es) be’ the study investigated some issues of global spread for the habitual marking with be and do(es) be and variants thereof on eWAVE: the electronic World Atlas of Varieties of English. This was felt necessary because the features studied in this paper repeatedly appeared in the scholarly literature comparative terms with other varieties of English across the globe. It was felt important to investigate how these IrE habitual features looked like on a global perspective.

eWave is an open access resource, an interactive database on morphosyntactic variation in spontaneous spoken English. These are differences in various Englishes, both on the morphological and syntactical level. 235 features from a dozen domains of grammar in 50 varieties of English are mapped out; traditional dialects, high-contact mother-tongue Englishes, indigenized second-language Englishes, as well as 26 English-based Pidgins and Creoles in eight Anglophone world regions. The data was available in an interface and several straightforward methods were available for extracting and analysing data on various features. World Geodetic System coordinates were given for each variety. On the negative side, the data could not be saved automatically however.

The eWave homepage showed tabs; varieties, features, informants, examples and sources, and these were explored further, the site being fairly user-friendly. Features of the various Englishes were listed below self-explanatory maps. Detailed columns listed further links to maps and the particular instances of the feature in each variety. Data was extracted for IrE on all habitual markers as well as comparative data covering these habitual features on a world wide scale. The information extracted is reported on the chapter dealing with results.

3.2 WebCorp Linguist's Search Engine

Searches also included corpora covered by the on-line Birmingham University Web Corpus (WebCorp), the three following alternatives searches were screened for IrE habitual markers.


6 (Available online at http://ewave-atlas.org/)
b. A Diachronic English Web Corpus (DIA): 130 million word corpus randomly selected from a larger collection.

c. Birmingham Blog Corpus: 630 million word corpus built from blogging websites.

The goal was to determine whether IrE specific habitual markers were to be found anywhere on the internet sites trawled by the WebCorp, Birmingham City University’s own search engine. The crawl process is illustrated in figure 4a below.

**Figure 4a. Process carried out on the WebCorp project.**

A request for data was sent from the user computer, and this was translated into the kind of search that a search engine can process. The engine then crawls internet sites and forums and sends back a list of URLs which WebCorp pulls in directly and returns to the interface in variously organised results, primarily in terms of concordance of the inputted request word or phrase. This process is a time-consuming task but Webcorp automates this trawling and scraping system. Returned hits for any word or phrase, on any of the corpora, were viewed together with the returned format (HTML for instance), short URL, number, date or domain, the latter being the general subject area in which the date was found (sports, entertainment, etc). Clicking on individual tokens will bring up the item in context, and links. Further options are available including Parts of Speech (POS) tags for identifying the composite parts of the returned language. Further discussion of data extracted as well as limitations encountered, will be presented in the results chapter below.
3.4 A Corpus of Global Web-based English, GloWbE

A Corpus of Global Web-based English, GloWbE is a corpus of 1.9 billion words from 1.8 million web pages in 20 different Anglophone countries. Data was collected from the (entire island of) Ireland section, just over one million words. The returned tokens were presented alternatively as a graph of tokens per country where the height of each column was proportional to the number of hits or returned tokens, or alternatively individual lists were presented from each country corpus. It was possible to draft comparisons between language features/words from two or more countries. Since this corpus was released in April 2013 it was expected that hits could be found for all habitual IrE markings dealt with in this paper, especially since a great amount of data originated from forums and social media websites. Searches for all habitual markers were executed as well as comparisons done for ‘do(es) be’ between the UK and Irish corpora.

3.5 Advanced web-based search or crawl of .ie social media

Evidence mounted in the literature that certain forms of the habitual marker in IrE could be very rare in corpora made up of standard Irish English. Ronan (2011) found habitual markers to be as low as one per million words i.e. only one habitual marker of the do(es) be category in the entire corpus of one million words (International Corpus of English, ICE Ireland). It became necessary to acquire data in a different manner to that of a standard corpus. Research began as to what was the most efficient method of gaining access to Internet forums in order to be able to search for instances of habitual markers in IrE. Difficulties in using internet data for linguistic research has been pointed out. Kilgarriff, A and G. Grefenstette (2003: 342) argue the limitations of the web asking whether it is not a ‘dirty corpus’ due to the amount of typos and unedited language. The search engine google too has its limitations: priority given to headlines in documents or websites means that documents or in-site text do not make up an equal share of all the hits that the search engine presents. Given that what is extracted and presented are first and foremost sites that are linked to i.e. ‘popular’ sites, any non-linked sites will not be searched.

It seemed likely however that habitual be hits in IrE might be revealed if selected sites or forums were targeted for a crawl from an advanced application program interface (ATP).  

---

7 I was familiar with the on-line Time Magazine Corpus material which is also created by Mark Davies, Professor of Linguistics at Brigham Young University.
8 Which in simple terms is the elaborate language used to conduct all manner of programming at the level of an application on say a device such as a computer or mobile phone.
Such requests are the communicative technology behind social media forums such as the microblog service Twitter (Crystal 2011:39-41). More numerous hits of ‘bes’ was revealed to be the case when rollercoaster.ie was targeted, for instance, in an advanced Google search\(^9\) on `site:https://www.rollercoaster.ie "bes"`. rollercoaster.ie is a forum dedicated to pregnancy and parenting with registered forum members from across Ireland. The following was returned for a young teenage female in a small town in Donegal, in the North West of Ireland:

(29) I hope for the child's sake that some more investigating bes done in this case.

The method was repeated for the following sites; Facebook.com, rollercoaster.ie, and the boards.ie forum, considered one of the largest indigenous websites in Ireland. Data was collected manually, instances counted and links followed to determined gender, class, etc. The estimation about social status was determined according to information that members had submitted on gender, job and location. This information was entered into a comparative table (see appendix) listing instances of habitual markers and source information.

The final resource that was determined to be a good source for collecting data was Twitter. Twitter is a free microblogging site for sharing information (via micro posts), photos and links.\(^{10}\) Twitter is perhaps one of the most interesting applications for linguistic research because a semi synchronic access to language is available at a limited amount of characters. The 140-character text messages are called "tweets". Complying with Twitter’s terms of service only excerpts or summaries of statistics can be published and no raw data can be shared (cf. Twitter’s Terms of Service).

Twitter advanced API searches (the process of which is similar to the one detailed above for the Google search) were conducted on all available feeds in the region of Ireland back to Twitter’s beginning in July 2006 to April 14\(^{\text{th}}\) 2014. A search address bar in the web browser could look like the following `"she bes" -RT lang:en near:"Ireland"`. The –RT was inputted when retweets were to be excluded from the search. A retweet is when a user quotes

---

\(^9\) Search conducted with google.ie using Boolean operators such as: exclusion "-xx", alternatives "xx OR yy" quotations marks (") for a phrase. A particular site or domain was targeted by including ‘site:’ in the search followed by the specific site address ‘site:http://www.facebook.com’ for instance. The latter searched with "bes" only returned one genuine habitual bes hit out of a great many others, some of which, like the taxation acronym ‘BES’, have been discussed above. Boolean operators could not rule out ‘BES’ for instance since that would also remove all instances of habitual bes.

\(^{10}\) There is a search or feed available called Twitter’s “garden hose” an API where access can be gained to a 10% feed of all the twitter messages which are sent at any one time. Details about this research were addressed to the research department at Twitter where these higher levels of access can be attained but no answer was received.
or ‘forwards’ another user on the application. In this way many users acquire followers who perhaps retweet them.

The inputted search was sometimes ‘she bes’ or ‘he bes’ since searches for the word ‘bes’ alone brought in huge volumes of ‘bad’ hits with typos and dropped ‘t’ from ‘best’ etc. By homing in on specific ordinance locations on the API search mechanism instances were traced to specific geographic locations. Users usually had some personal information available. Searches were confined to parameters around Dublin, Belfast, Dundalk, Derry, Cork, etc. Entering these parameters even returned global Tweets regardless. Anonymity was adhered to in order that there be no infringement of a person’s rights under the applicable law\(^\text{11}\). The information data retrieved is freely available to the public however. Whether or not the information made available is true or not is a matter of speculation.

4. 0 RESULTS

4.1 eWAVE The Electronic World Atlas for Varieties of English: data and results

Questions looked at on eWAVE were the following:

i. Were habitual markers, be(s) and do(es) be, (feature X in this case) represented in the Irish English (IrE) variety, and how widespread were these habitual markers in Englishes worldwide?

ii. Were certain habitual markers particular to Irish English or not? How many varieties of English had that particular feature? Was this a worldwide feature?

eWAVE presented the following information for what it termed ‘invariant be’.

\(^{11}\) (cf. European Data Protection Directive).
23

Datapoint Irish English/Invariant be as habitual marker

Variety: Irish English
Feature: Invariant be as habitual marker
Value: B - feature is neither pervasive nor extremely rare
Informants: Markku Filppula

Example 3975: It's better, because you**be's** bored doing nothing at home

<table>
<thead>
<tr>
<th>Feature status</th>
<th>Englises</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="red-circle.png" alt="Red circle" /> - feature is pervasive or obligatory</td>
<td>2</td>
</tr>
<tr>
<td><img src="orange-circle.png" alt="Orange circle" /> - feature is neither pervasive nor extremely rare</td>
<td>8</td>
</tr>
<tr>
<td><img src="green-circle.png" alt="Green circle" /> - feature exists, but is extremely rare</td>
<td>7</td>
</tr>
<tr>
<td><img src="yellow-circle.png" alt="Yellow circle" /> - attested absence of feature</td>
<td>50</td>
</tr>
<tr>
<td><img src="empty-circle.png" alt="Empty circle" /> - feature is not applicable (given the structural make-up of the variety/P/C)</td>
<td>8</td>
</tr>
<tr>
<td><img src="question-mark.png" alt="Question mark" /> - no information on feature is available</td>
<td>1</td>
</tr>
</tbody>
</table>

Fig. 3a. Invariant be as habitual marker

In addition to Irish English those included in category B above (feature is neither pervasive nor extremely rare) was Bahamian Creole, Bahamian English, Butler English, Indian South African English, Rural African American Vernacular English, Tristan da Cunha English and Vernacular Liberian English.

Datapoint Irish English/Do as habitual marker

Variety: Irish English
Feature: **Do as habitual marker**
Value: B - feature is neither pervasive nor extremely rare
Informants: Markku Filppula

Example 3976: Two lorrys of them [turf] now in the year we do burn.

<table>
<thead>
<tr>
<th>Feature status</th>
<th>Englises</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="red-circle.png" alt="Red circle" /> - feature is pervasive or obligatory</td>
<td>5</td>
</tr>
<tr>
<td><img src="orange-circle.png" alt="Orange circle" /> - feature is neither pervasive nor extremely rare</td>
<td>5</td>
</tr>
<tr>
<td><img src="green-circle.png" alt="Green circle" /> - feature exists, but is extremely rare</td>
<td>12</td>
</tr>
<tr>
<td><img src="yellow-circle.png" alt="Yellow circle" /> - attested absence of feature</td>
<td>45</td>
</tr>
<tr>
<td><img src="empty-circle.png" alt="Empty circle" /> - feature is not applicable (given the structural make-up of the variety/P/C)</td>
<td>6</td>
</tr>
<tr>
<td><img src="question-mark.png" alt="Question mark" /> - no information on feature is available</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 3b. Do as habitual marker.

---

12 Butler English is structurally akin to pidgin and developed in India among the servant classes who waited in the houses of the British Raj.
13 Tristan da Cunha English is the language spoken in the “the most remote inhabited archipelago in the world, lying 2,000 kilometers (1,243 mi) from nearest inhabited land, Saint Helena, 2,400 kilometres (1,491 mi) from the nearest continental land, South Africa, and 3,360 kilometers (2,088 mi) from South America.” (Wikipedia).
Included in category B above (feature is neither pervasive nor extremely rare) is Irish English, Bahamian, Pakistani English, St. Helena English and Welsh English.

The higher instance of category A (feature is pervasive or obligatory) included Barbadian Creole (Bajan), Gullah and Guyanese Creole (Creolese).

Datapoint Irish English/Other non-standard habitual markers: analytic

| Variety: | Irish English |
| Feature: | Other non-standard habitual markers: analytic |
| Value: | B - feature is neither pervasive nor extremely rare |
| Informants: | Markku Filppula |

Example 3977: They do be shooting there couple of times a week or so.

<table>
<thead>
<tr>
<th>Feature status</th>
<th>Engishes</th>
</tr>
</thead>
<tbody>
<tr>
<td>A - feature is pervasive or obligatory</td>
<td>13</td>
</tr>
<tr>
<td>B - feature is neither pervasive nor extremely rare</td>
<td>8</td>
</tr>
<tr>
<td>C - feature exists, but is extremely rare</td>
<td>11</td>
</tr>
<tr>
<td>D - attested absence of feature</td>
<td>35</td>
</tr>
<tr>
<td>X - feature is not applicable (given the structural make-up of the variety/P/C)</td>
<td>5</td>
</tr>
<tr>
<td>? - no information on feature is available</td>
<td>4</td>
</tr>
</tbody>
</table>

Figure 3 c. Feature status, other non-standard habitual markers.

So The Electronic World Atlas for Varieties of English presented these three features of habitual aspect as representative of Irish English.

a) Invariant be as habitual marker
b) Do as habitual marker
c) Other non-standard habitual markers

We have encountered these in the literature, in one form or another; ‘be’s’, ‘do’, ‘do be’. And many of these features were seen to be widespread globally. It is interesting to see that since it has been hypothesized that invariant be (and be/bes) to express habitual aspect, realisations of perfect aspect, and wider use of the progressive, are also assumed to be IrE influences on other varieties of English. The major direction points towards the islands of the Caribbean (Bahamian) and the southwest coast of North America (Gullah). It is beyond the scope of this paper however to comment on the matter further but relationships between many habitual markers in IrE and many creoles are remarkably striking.
4.2 Results from the WebCorp Linguist’s Search Engine data

Some of the problems encountered with a web based corpus search, or data extracted on a crawl of web-based forums can be immediacy taken up here by way of introduction. While the WebCorp project is linked to the UK city University the actual data returned from a search (of ‘does be’ in this instance) is not location specific and neither is any geographical information listed in the returned information. Data must be analyzed by following the actual HTML links to determine the possible location, gender or social status of a speaker.

A search of ‘does be’ on the Birmingham Blog returned 18 instances in 3 min 28 seconds (Fig 1 in appendix 1a). But further analysis determined that only one instance could be considered an actual IrE habitual marker. This was an anonymous person, ‘MD’, commenting on a blog the subject of which was the British television show, the X factor.

(30) …she wouldn’t deserve a tenth of the stuff that does be said about her!

Although we might speculate that this person could well be Irish or of Irish decent, and that they are commenting on a British television show which is available in Ireland, the hypotheses cannot be settled however. Initially data extracted from the Birmingham University Webcorp seemed impressive but this turned out to be somewhat superficial in character as further analysis revealed that there was only a quasi-calculative return in most instances. This superficial return of data was due to the lack of true hits for habitual markers and lack of information concerning the location of users. The Birmingham blog element of the web corpus returned only the 4 other hits below for does be.

(31) hey everyone,. i does be back lunch was nice
(32) dish it and everyone does be able to take it.
(33) Joba does be throwin' those sliders too
(34) the stuff that does be said about her!

Out of in 18 hits returned for the Birmingham blog element only these 5 tokens were found to be true ‘does be’ habitual markers. There were many repetitions that the engine has not processed in any manner but just presented them in the returns. The following hit is an example of an obvious typo (the ‘be’ should be ‘he’) given the context and followed up on (the site accessed manually).
Quick thought…who does be bow to now? The terrorists he seems to
coddle, or the Jews who help him get elected…”

The synchronic element SYN (fig 2 in appendix) returned almost all instances where the source text, although obviously IrE, was about a century old; the writers Joyce, Yeats, etc, or was dated to a large degree. In contrast, and paradoxically, although instances of habitual markers were fewer in the Diachronic corpus, DIA, (see fig 3 in appendix) had more contemporary tokens. The opposite was to have been expected since a diachronic corpus ought to show instances over a wider span of years and synchronic ought to present a contemporized section of hits only. The search for the word “bes” returned an extremely large number of typos of the word best; omissions of the ‘t’ generally and even an Afro American Vernacular dropping of the ‘t’ in ‘best’ as a feature of that English. There were no instances of habitual markers using ‘bes’, ‘be’s’ or ‘bees’ in any of the WebCorp elements.

A comparison was executed between returned WebCorp hits on ‘does be’ for diachronic (DIA) and synchronous English Web Corpus searches (SYN) (see figure 4a below). The general web addresses of the sources are given in columns. The column on the left is the DIA search and the column on the right is the web addresses of the SYN search.

<table>
<thead>
<tr>
<th>DIA/search</th>
<th>Result</th>
<th>Frequency</th>
<th>Number of Documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>✔ does be</td>
<td>does(VVZ) be(VB)</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>donaldcorrell.com (synge)</td>
<td>solargeneral.com</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>forum.rpg.net</td>
<td>robotwisdom.com</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>moneysavingexpert.com</td>
<td>authorama.com</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>redress.btinternet.co.uk</td>
<td>americanliterature.com</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>members.shaw.ca</td>
<td>mysite.verizon.net</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>conservativehome.blogs.com</td>
<td>schmolotics.blogspot.com</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SYN/search</th>
<th>Result</th>
<th>Frequency</th>
<th>Number of Documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>✔ does be</td>
<td>does(VVZ) be(VB)</td>
<td>8</td>
<td>6+2pdf</td>
</tr>
</tbody>
</table>

Figure 4a. Comparison of returned hits on ‘does be’ between diachronic and synchronous English Web Corpus using the WebCorp search engine.

14right***bright on May 17, 2010 at 8:54 AM' http://hotair.com/archives/2010/05/16/rahm-we-screwed-up-the-messaging-on-israel/
The table below shows frequency of results over time. A frequency per million words in this case, was also extracted for the synchronic corpus. Raw frequency was also determined. Figure 4 b shows the frequency of results, plotted for each month, over time. The blue dashed line shows the raw frequency value whilst the constant blue line shows the moving average.

![Graph showing frequency of results, raw frequency and moving average for 'does be'.](image)

In summary, the on-line WebCorp Linguist's Search Engine out of Birmingham University was not very revealing about forms of habitual be in IrE but it presented more hits for 'does be' in all three corpora used. The frequency was low but expectantly so given that the majority of the texts drawn from the internet are established or standard Irish English texts, many turned out to be a century or more old. The fundamental lack of hits for habitual 'be(s) was also a factor in furthering the search to the next corpus on the list since it was my opinion that more data could be extracted for the be(s) form at least. The mathematics of the frequency table above seem fairly well anchored but sadly the table only reflects the low frequencies seen for 'do(es) be' and cannot be truly determined in the geographical sense.

### 4.3 Results from the GloWbE data

There were extremely low hits for habitual features of the bes form in the GloWbE. (A great many BES were found since this is an investment company as well as a so-called Business Expansion Scheme, a tax incentive, in the republic of Ireland). Only one true token was found
in fact. Of the two that were presented one, a sentence from James Joyce, was a typo (34) for ‘he’. Although spelt ‘be’s’ on the internet text, a quick look at the printed work reveals the typo. (Several instances of ‘does be’ are actually found in Joyce’s *Dubliners*).

(36) What about him? # -- Well, poor fellow, be’s a decent sort of chap, after all, continued Gabriel in a false voice.

There was one hit for ‘be’s’ with an apostrophe. The source was also a dated story called The Friars of Urlaur.

A comparison of the hits for ‘do(es) be’ and were executed for the sections of the Great Britain corpus and the ‘Ireland’ corpus (fig 5 a below). This revealed that although there were many more hits for ‘do be’ in the GB section than the IRE section the GB hits were entirely non-habitual and the IRE hits were almost all habitual markers. The GB hits were almost entirely made up of emphatic or contrastive ‘do’ which is a standard grammatical feature found in most Englishes. An example of an emphatic do is the following:

(37) If you decide to use paint or permanent markers, do be very careful not mark the wood of your harp. (14 IEG)

There were no GB returns for ‘does be’ at all in the GB section. The number of words for the GB and IRE corpora were 387,615,074 and 101,029,231 words respectively. Token number denote raw frequency and PM stands for per million. Ratio stands for the ratio of one word to the other. i.e. do:be.

**IRE SEC 1: 101,029,231 WORDS**

<table>
<thead>
<tr>
<th>WORD/PHRASE</th>
<th>TOKENS 1</th>
<th>TOKENS 2</th>
<th>PM 1</th>
<th>PM 2</th>
<th>RATIO</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1</strong> DOES BE</td>
<td>21</td>
<td>7</td>
<td>0.21</td>
<td>0.02</td>
<td>11.51</td>
</tr>
<tr>
<td><strong>2</strong> DO BE</td>
<td>61</td>
<td>213</td>
<td>0.60</td>
<td>0.55</td>
<td>1.10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>WORD/PHRASE</th>
<th>TOKENS 2</th>
<th>TOKENS 1</th>
<th>PM 2</th>
<th>PM 1</th>
<th>RATIO</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2</strong> DO BE</td>
<td>213</td>
<td>61</td>
<td>0.55</td>
<td>0.60</td>
<td>0.91</td>
</tr>
</tbody>
</table>

**GB SEC 2: 387,615,074 WORDS**

Figure 5a. Comparison between hits on the GB section and the IRL section for do(es) be.
Out of 21 tokens or hits received for ‘does be’ in the IRL section (see fig 3 in appendix) 16 were habitual markers. The remainder, if not doubles, i.e. the same hit reported twice (perhaps on two separate sources or forums). The internet links can be followed, in most cases, and the gender of the user identified in many cases. Number 2 in the appendix below is obviously a woman given that she refers to her husband in the actual sentence. Furthermore she is at the lower spectrum of the economic scene since her husband gets ‘dole’, the vernacular term in the republic of Ireland for receiving unemployment benefit. Although most speakers were working class there were exceptions, such as the boards.ie hit (No. 4) where a BSc in Environmental Science and the issue of unemployment in that field implied that the phenomenon of ‘does be’ was not solely restricted to an uneducated working class. The site called digitaldarragh.com was not open to the public so no further data could be collected. A simple internet search of number 4 revealed that although the link to the web forum in the corpus was broken it was possible to trace the speaker (a fisherman in this instance). 6 and 7 are indeed older than 50 years. Although the web pages of Rollercoaster.ie is down permanently, by proceeding to a search of google in .ie, the Irish top level domain (ccTLD), and inputting a section of the phrase in hit 10, it was discovered that the phrase was again from a novel of Joyce. The same procedure revealed 13 to be a woman (the Coombe is a Dublin maternity hospital) with severe spelling problems and restricted education. The remainder were revealed to be mostly male but 14 were female.

The search for habitual be(s) however was again a non-starter since there were only two hits and one of these was an obvious typo as previously stated.

4.4 Results from advanced web-based searches of .ie social media

Considering the distinct lack of data on be(s) or any other similar habitual be in all the sources explored previously I was of the opinion that the search for empirical instances of the habitual be(s) would and could be revealed on social forum sites and other internet media. I myself was born in Dundalk on the border of Northern Ireland and have personally witnessed all habitual features mentioned in this paper, especially the ‘bes’ form, to have been common features of IrE in that region at least. Instances of ‘bes’ did occur at a few of the forums which were sources for the extracted data of the GloWbE corpus, even though when searches for ‘bes’ were performed in GloWbE no such data was revealed for ‘bes’. In order not to drop the issue prematurely and simply conclude that evidence suggested that the habitual be had become discontinued in modern IrE advanced web searches were executed in the manner and method described in chapter 3. The following data was retrieved using the methods described there.
Habitual bes was returned in a good many cases of searches such as site:https://www.rollercoaster.ie "bes". [R]ollercoster.ie is a forum dedicated to pregnancy and parenting.

(38) She bes delighted with herself
(39) the time i do fall asleep ds bes wakening
(40) Oh god my mam says "he bes". I only say "I do be" if I'm taking the piss. That and "I seen him"

37 above is an infrequent example of habitual do and bes in the sentence phrase, as is 38.

<table>
<thead>
<tr>
<th>WORD/PHRASE</th>
<th>PHRASE</th>
<th>MALE / FEMALE</th>
<th>HITS</th>
<th>AGES</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 rollercoaster.ie does be F, (F), (F), (F) 4 25,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 rollercoaster.ie do be F 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 facebook .com bes F 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 facebook.com does be 5 F, 5 M 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 boards.ie bes 2F, 6M 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 6 a. Instances of do(es) be and bes in 3 forums (2 Irish).

Chapter 3 above has explained the method behind the searches and services available for the Microblog Twitter. The 140-character text messages called "tweets" were extracted by hand and the summary of the data extracted for habitual be(s) here, in particular, is by no means in contravention of Twitter’s Terms of Service under sharing and publication of raw data. Specific ordinance locations on the API search revealed instances which could be traced to specific external sites where the personal information of users has been freely available to the public. Again, anonymity was adhered to in order that there be no infringement of a person’s rights under EU law.

It was beyond the scope and resources of this paper to be able to achieve any frequencies on the Twitter data. In theory this could be possible to attain given the limited nature of the characters allowed in a ‘tweet’ (140 characters). Table 6b in the appendix lists examples of Tweets with habitual bes on the Twitter microblog application. A good number of instances of habitual be(s) were presented in the Twitter data considering the longstanding lack of instances of be(s) available in all the first sources, GloWbE, WebCorp, etc. The largest proportion of locations traceable for ‘be(s)’ was in the north of the country,
particularly along the border region of Northern Ireland; Dundalk, Co Louth for instance, had many instances (fig a below).

Donegal in the extreme North West of Ireland had several instances of ‘be(s)’ as did Derry, and other border regions and finally Belfast. This is in line with the existing scholarship on the subject and the phenomenon has been commented on by Kallen (1994:180), who has mainly associated this form with the Ulster dialect in the North. Donegal and Derry as well as Dundalk in North Co Louth, constitute locations in Ireland that had been almost entirely made up of the Native Irish speaking population, regions of Donegal remain Irish speaking communities today.

This proximity to Irish, if not in physical and terms, than perhaps in terms of time (Irish being spoken as a native language in the not too distant past or else still being spoken) might
seems to be the element that contributes to the vitality of the ‘bes’ form in the North and around the border as Hickey and others maintain. But Hickey (2007:231) points out too that bes is not only confined to the region of the North and border of Ulster but is found in the south too, in places such as Wexford in the extreme South East. This study found no evidence to support Hickey on that point. But there were several instances of the ‘bes’ form in Dublin, but these could be from people who had moved to the city from the areas mentioned above.

Like Shimida (2013:255) it may be conjectured that awareness of bad grammar deters people from using a parochial dialect in ‘traditional’ public settings at least. This may be one of the reasons why there is a lack of the data in the years following the era when Irish writers such as Joyce, Synge and Yeats were churning out vernacular IrE in published form. From that period, of almost a century ago, until the modern age of IT and the World Wide Web few instances of habitual markers seem to be available on corpora largely dealing with web searches and crawls. This study shows that there has been somewhat of a revival of the vernacular habitual markers given the surge of the population who had taken to using internet forums and applications such as the mini-blog Twitter and the social networking service, Facebook. Common use of the vernacular in such relaxed settings as internet forums on parenting for instance, have upped the tallies of habitual markers in IrE, even if these features of the language have a tendency to die out given the overwhelming availability of Standard English (2013: 2).

5.0 CONCLUSION

The questions asked the beginning of this paper were the following:

a. How are habitual aspect markers used in IrE?
b. In what linguistic contexts is the habitual aspect realised primarily?
c. How common are habitual aspect markers in a specific corpus or internet forum?
d. What type of speakers use these forms primarily?

The answer to question ‘a’ is that the markers, do(es) be and be(s), although they are represented in older texts (Synge, Joyce, Yeats, etc. in ) are currently a feature of IrE on web forums and in so-called social media applications such as Twitter and Facebook. The markers are used mainly in an informal sense and in a relaxed setting where people interact in an instructive but social manner.

In terms of how common these features are (question c) it has been demonstrated that frequencies of true habitual markers with be(s) and do(es) be in corpora were low but
applicable instances were measured in the region of 7 in one million words: a ratio if
approximately 1 per 142,857 words for ‘does be’ in the extracted data of the GloWbE corpus for
Ireland. No extracted frequency could be determined for be(s) or any other similar habitual be
variant in this paper other than that of one 1:1 000 000 for the one instance that habitual be(s)
was discovered in GloWbE.

It is difficult to calculate the Twitter or Facebook frequencies of be(s) or in any of the
sources explored for be(s) in this paper, but a very open estimate of 7 tweets per year for ‘bes’
can be reconciled with just over 15 ‘bes’ tweets between February 2012 and April 2014. (fig.
6b above) 17 instances of habitual be(e) tokens were counted for Ireland but two of these can
be attributed to England. The average number of tweets per day for Ireland can only be a
matter of speculation since large corporations have invested interests in publishing wildly
speculative statistics. In conclusion, the be(s) feature is ‘neither pervasive nor extremely rare,
to quote the terminology of the Electronic World Atlas for Varieties of English. The type of
speaker using the habitual be is in general terms the Working class male and female, but the
feature is not entirely ruled out of middle class circles (the forum on higher level education
for instance has pointed to this (p. 29-30). The data extracted for this study has uncovered,
however, that poorly educated people, at least those who are bad at spelling, fall into the
majority of the users of all habitual markers.

The frequencies of all habitual markers mentioned are incredibly low in corpora
composed of standardised Irish English texts, which is the case for the main part of all the
corpora analysed in this study -except for the targeted forums, Twitter and Facebook (see fig
7 below). Numerous attempts were made to gain access to spoken corpora of Irish English as
mentioned in chapter 3. I requested access to these on many occasions, but no access was
gained. Using the only methods at my disposal, the analysis of this paper reveals that
applications such as the Facebook (SNS) and the mini blog Twitter can be exploited as major
sources for accessing habitual markers in IrE.

<table>
<thead>
<tr>
<th>Corpus</th>
<th>does be</th>
<th>bes/bees/be's</th>
<th>do be</th>
<th>usually</th>
<th>Per mil</th>
</tr>
</thead>
<tbody>
<tr>
<td>WebCorp DIA</td>
<td>7</td>
<td>0</td>
<td></td>
<td>1654</td>
<td></td>
</tr>
<tr>
<td>WebCorp SYN</td>
<td>8</td>
<td>0</td>
<td></td>
<td>55650</td>
<td></td>
</tr>
<tr>
<td>Birmingham blog</td>
<td>5</td>
<td>0</td>
<td></td>
<td>81104</td>
<td></td>
</tr>
<tr>
<td>GloWbE (Ire)</td>
<td>16</td>
<td>1</td>
<td>26</td>
<td>18481</td>
<td>0.16/0.01/0.26</td>
</tr>
<tr>
<td>GloWbE (GB)</td>
<td>1</td>
<td>4</td>
<td></td>
<td>65974</td>
<td>0.10/.../0.40</td>
</tr>
<tr>
<td>Twitter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facebook</td>
<td>10 (5M/5F)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rollercoaster.ie</td>
<td>4</td>
<td></td>
<td>1</td>
<td>1497</td>
<td>0.20/150.25</td>
</tr>
<tr>
<td>BYU/BNC/spoken</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Frequencies in habitual markers across the corpora
Returns for the common habitual marker ‘usually’ in StE however (table 1 above) dwarf the frequencies for the habitual markers do(es) be and bes. Older data, almost a century old, incorporated digitally into major corpora such as GloWbE, gave false positives in terms of frequencies. The drama entitled the Playboy of the Western World, by J. M. Synge, has 7 instances of habitual ‘Do Be’ for instance. Uploaded onto the World Wide Web such literary works are a fundamental source for linguistic research, yes, but any corpus trawling the web for hits of Habitual ‘do be’, for example, will present these instances as occurring on the date when they have been last ‘edited’ on-line. Such a scenario hampers any attempt to find adequate synchronic data for a final linguistic analysis of the feature.

Indications are that the feature of habitual be(s) has been exported to England and London in particular. These instances were taken into account in the analysis since they were particularly interesting; they could be traced further by following the links to the person’s Twitter account and even Facebook in some instances. One Tweet from Figure 6b, for instance, indicates that a female, living in London and using habitual features in her Tweeting is mixing the common IrE habitual marker ‘bes’ with a form of be that is a standard feature of African American Vernacular English.

This first ‘be’ is an AAVE habitual be (Wolfram: 2000) and the second ‘bes’ is an IrE habitual marker. This person is Irish in all probability given that her surname is a typical Irish surname and that she is using the Irish bes form of the habitual marker which has not, to my knowledge, been reported in England by any other linguists studying the field.

The hypothesis stated at the beginning of this paper is borne out by the results from the analysis which have shown that the feature of habitual marking by be(es) and do(es) bes in IrE has not receded but is a vital element of IrE at the interface where people use the language in their daily lives. As predicted, the feature has very low frequencies in standardised Irish English, but new forums for communication have allowed speakers of IrE to drop their grammatical guard in the company of their peers and the habitual feature can be discerned in these small texts.

---

REFERENCES


