# Exploring the role of token frequency in phonological change: evidence from TH-Fronting in east-central Scotland[1]

LYNN CLARK

*University of Edinburgh*

GRAEME TROUSDALE

*University of Edinburgh*

ABSTRACT

Recent research on frequency effects in phonology suggest that word frequency is often a significant motivating factor in the spread of sound change through the lexicon. However, there is conflicting evidence regarding the exact nature of the relationship between phonological change and word frequency. This article investigates the role of lexical frequency in the spread of the well-known sound change TH-Fronting in an under-researched dialect area in east-central Scotland. Using data from a corpus of conversations compiled over a two year period by the first author, we explore how the process of TH-Fronting is complicated in this community by the existence of certain local variants which are lexically restricted, and we question to what extent the frequency patterns that are apparent in these data are consistent with generalisations made in the wider literature on the relationship between lexical frequency and phonological change.

## INTRODUCTION

This article is a contribution to the on-going debate regarding the role of frequency in phonological change, based on data collected from a community in east central Scotland by the first author. The phonological change we describe is the spread of TH-Fronting (Wells 1982) into the accents of speakers from a community in Fife. Although Wells invokes 'TH-Fronting' to refer to 'the replacement of the dental fricatives [θ,ð] with the labiodentals [f] and [v] respectively' (Wells 1982:328), we follow Stuart-Smith & Timmins (2006) who adopt the term only with reference to the voiceless variants. The first reported evidence of TH-Fronting in Scotland is given by Macafee (1983: 54) as occasional and sporadic but the main body of research on TH-Fronting in Scotland comes from the analysis of two corpora collected in 1997 and 2003, both of which form part of a much larger research project on language variation and change in Glasgow (Stuart-Smith & Tweedie 2000). The spread of TH-Fronting has also recently been investigated in the New Town of Livingston (Robinson 2005) which is situated between Edinburgh and Glasgow, approximately 15 miles to the west of the former and 30 miles to the east of the latter.

A brief overview of the sociolinguistic literature on TH-Fronting in non-standard varieties across the British Isles (e.g. Kerswill and Williams 1999; Kerswill 2003; Stuart-Smith and Timmins 2006; Robinson 2005) reveals a number of consistencies between the use of 'fronted' variants and the social factors age, gender and social class. The main finding seems to be that TH-Fronting is favoured by working class adolescent males in most communities. However, none of these projects has yet examined the effect of lexical frequency as a motivating factor in the spread of this sound change. This is perhaps surprising given that recent research on

sociophonetics and phonological change has prompted a considerable debate on the role of frequency in the propagation of innovations. We begin by reviewing some of this research and then, in what follows, we expand the analysis of variation and change in TH-Fronting in Scotland to explore the role of token frequency as a possible motivating factor in the spread of this innovation.

## 1. THE ROLE OF LEXICAL FREQUENCY IN SOUND CHANGE

Certain researchers (e.g. Bybee 2007, Philips 2006) have provided evidence in support of the claim that frequency has a very significant role to play in the spread of sound change, while others (e.g. Labov 2006) have provided other sets of evidence which has downplayed the importance of frequency. This apparently contradictory state of affairs may be in part due to the fact that some phonological changes (e.g. the spread of /ð/ deletion in Spanish, Bybee 2002) have been shown to affect high frequency words first, some others (e.g. the unrounding of high-mid rounded vowels in Middle English, Philips 1984) affect low frequency words first, and yet others (e.g. the spread of 'Canadian Raising' in present-day American English, Labov 2006) show no effects of frequency at all. In order to contextualise our findings within the larger discourse on frequency and change, we provide a summary of the contrasting positions taken on the role of frequency by Bybee (2007), Phillips (2006) and Labov (2006)[2] and discuss how the results of the correlations in our data compare with certain generalisations that have been proposed in the literature on frequency effects in phonological change.

### *1.1 Bybee (2007)*

Joan Bybee's work on frequency is wide-ranging, and so for present purposes, we focus only on a particular set of distinctions she has made regarding token frequency. We do not deal with type frequency at all in this article. In her discussion of token frequency effects, Bybee has noted two distinct tendencies, which she labels the Conserving Effect (hereafter CE) and the Reduction or Reducing Effect (hereafter RE). The CE suggests that frequent use of linguistic tokens strengthens the mental representation of those tokens; in the terminology of the usage-based model (e.g. Langacker 1987), such representations become more and more entrenched, with the result that they are more directly accessible (since they acquire unit status), and are more resistant to analogical changes. For instance, the regularisation of the past tense of *climb* from early Modern English *clomb*:

(1) So clomb this first grand Thief into Gods Fould (Milton [1667] *Paradise Lost* Book IV, 192)

to *climbed* in present-day English, stands in contrast to the past tense of *run*, i.e. *ran*, which shows that the verb has remained 'strong'. This can be explained as a result of the CE, since the higher frequency of tokens of *ran* compared to tokens of *clomb* protects the former from undergoing analogical change as rapidly as the latter. By contrast, the RE suggests that high frequency tokens regularly undergo attrition, whether this be part of a process of lexicalization from a phrase to a word (e.g. *How do you do* > *Howdy*) or a process of lenition within a particular morpheme (e.g. *t/d* deletion in *last week*). Bybee has argued that the RE occurs because "repetition of neuromotor sequences leads to greater overlap and reduction of the component articulatory gestures" (Bybee 2007: 11). The RE therefore means that high frequency tokens are more readily affected by reductive sound changes than low frequency tokens are.

*1.2 Phillips' (2006)*

Phillips invokes the concept of 'lexical analysis' to explain her interpretation of the relationship between sound change and lexical frequency: "changes which require analysis…during their implementation affect the least frequent words first, others affect the most frequent words first" (2006: 56). This hypothesis is the consequence of an investigation in Phillips (1998) on changing stress shift patterns in verbs with the –*ate* suffix in English. The study in question examined the relationship between lexical frequency and the stress placement in verbs like *lactate, pulsate* and *stagnate* (where the stress placement is variable but typically initial) compared with verbs like *frustrate* and *dictate* (where the stress is final). Phillips (2006:41) explains that this stress shift has been in progress for over a century. This sound change is not physiologically motivated and yet it follows a pattern often associated with reduction and assimilation processes because high frequency words are changing first.

The idea expressed in Philips (2006) is that sound changes which affect only the phonetic realisation of lexical items, without first invoking the abstract generalisations (or schemas) that have emerged from these word forms in the grammar, affect the most frequently used words first. These changes are typically (although not always) physiologically motivated changes such as assimilations and reductions. Changes which require access to a deeper level of lexical representation such as phonotactic constraints or generalisations over stress patterns affect the least frequent words in the language first. These changes typically involve analogical levelling.

Explanations for the generalisations proposed by Phillips (2006) are based on the discussion of lexical analysis in Bybee (1985: 118). Here, Bybee claims that lexical items which have high token frequency are less dependent on their related base words, are more autonomous and therefore less likely to undergo 'analysis' than items

with low token frequency which, by contrast, tend to require more access to information in the 'lexical entry' including generalisations drawn on the part of the speaker across related or similar items in the grammar.

### *1.3 Labov (2006)*

Labov discusses frequency within a larger context of sociophonetics and language change, and suggests that data from the *Atlas of North American English* (Labov et al. 2006) do not support some of the predictions made by exemplar theory regarding the role of frequency in the spread of phonological innovations[3]. For instance, in Columbus, Ohio, the fronting of /uw/ in morphemes without a following liquid has no correlation with frequency – there is no difference in the likelihood of fronting with highly frequent *do* when compared with the more infrequent *dew*. Instead, the spread of /uw/-fronting is primarily conditioned by linguistic environment, with the progression of change inhibited in morphemes in which a liquid follows (Labov 2006: 509-510). This is not to say that token frequency may never have an effect in the spread of this change: rather, "as the change progresses, it is still dominated by phonetic factors, but within these constraints, the variation can show small lexical as well as social effects" (Labov 2006: 511). Thus the relative weighting of frequency with regard to other factors in change is fundamental to Labov's analysis. This point is central to the remainder of our discussion and we deal with it specifically in relation to the data from east-central Scotland in the following section.

## 2. METHODS

### *2.1 Collecting the data*

The data presented here were collected from a group of 54 speakers who play together in two interrelated pipe bands that will be known as West Fife High Pipe Band (hereafter WFHPB) in west Fife, the area shown in Figure 1.

[Figure 1 about here]

The area in which WFHPB are based is located around 22 miles north of Edinburgh, over the Firth of Forth. The data were collected by the first author over a period of 30 months using the ethnographic technique of long-term participant observation (Eckert 2000). The conversations that comprise the majority of the corpus were collected in the summer of 2006 and centre on a sorting task that the informants were asked to complete in small groups of friends[4]. The resulting data consists of 38 hours of recorded speech which have been fully transcribed and amounts to a corpus of 360,000 words. All instances of the variable (th) were extracted from the corpus and then random selections of these tokens were cross-checked for accuracy of transcription. This resulted in an initial data pool of 5205 sites of (th).

*2.2 Measuring lexical frequency*

In order to discover whether there is a significant correlation between lexical frequency and the role of TH-Fronting in WFHPB, it is first necessary to consider how best to measure lexical frequency as there are a variety of different methods available. We consider frequency of use as a local phenomenon and so measure the lexical frequency of a particular item against the frequency of other items only in this locally-based corpus. Often researchers interested in frequency effects take the frequency value of a particular lexical item from a large corpus such as the Brown

Corpus (e.g. Dinking 2007 and Abramowicz 2006) or from a list of frequency counts such as that provided by Baayen et al. (1995) in the form of the CELEX lexical database (employed by Hay 2001). However, certain local forms (e.g. place names, nicknames and other non-standard lexical items) which occur fairly frequently in the WFHPB corpus are much less frequent in a 100 million word corpus of British English such as the BNC or even a more local corpus of Scottish English such as the SCOTS corpus. This was a particular problem for our analysis because a large amount of the conversations that took place were in non-standard dialect and so had the frequency counts for these non-standard lexical items come from a large database of lexical frequency rather than the WFHPB corpus itself, the frequency value assigned to these items would not have been an accurate representation of the frequency with which they are used by these speakers. We also decided (where possible) to avoid categorising lexical frequency into discrete categories such as 'high frequency' and 'low frequency'. Instead we follow Hay (2001) in treating frequency as a gradient phenomenon. We therefore assume that lexical frequency is gradual and relative, not categorical or universal.

Most frequency research to date has examined the effects of lexical frequency on variation and change in isolation and so we begin our analysis by following this typical procedure. A Pearson's correlation was initially used to measures the extent to which values on the variables 'lexical frequency' and '(th): [f]' co-vary. Also, following Hay & Baayen (2002), the measurements of lexical frequency and the token frequency of (th): [f] were converted into a logarithmic transformation since "there is evidence that humans process frequency information in a logarithmic manner – with differences amongst lower frequencies appearing more salient than equivalent differences amongst higher frequencies" (2002: 208). The data were normalized

using the Log10 transformation. This transformation was selected because in its raw form, the data has a moderate positive skew based on the analysis of kurtosis, skewness and Kolmogorov-Smornov Z test of normality (see de Vaus 2002: ch11 for details of these tests for normality). Also, because some of the raw frequency data for (th): [f] contains a value of 0, and there is no logarithm of the value 0, it is necessary to add a constant to the original values in the transformation. In this case, we simply added the value of 1 to the raw frequency scores.

The lack of a standard method to conduct correlations on frequency research led us at the outset to correlate word frequency with the number of tokens of a particular word in which the variable is realised with one variant (in this case, the number of tokens of (th) realised as [f]). However, as Kapatsinski (pc.) points out, it is possible that these two variables may correlate independently of any frequency effect using this method. In order to combat potential interference, it was therefore necessary to instead correlate word frequency with frequency of (th): [f]/word frequency. In other words, this method correlates the proportion of each word in the corpus which appears with (th): [f] against the lexical frequency of that word. While this may be a more accurate method of calculating lexical frequency in a large corpus, this method may be less well suited to a smaller corpus such as the WFHPB corpus or to low-frequency lexical items (only those lexical items with 3 or more tokens were included in the analysis) because it depends on large numbers for accurate results. This method of correlating lexical frequency with phonological change is therefore only likely to find a significant result if the effect of lexical frequency is very large. The results of this correlation are charted in figure 2.

[figure 2 about here]

Notice that there are a large number of lexical items, some of which have very high frequency counts, which are not participating in this change to [f]. We shall deal with these lexical exceptions first before moving on to analyse the remainder of the data.

*2.3 Exceptions to TH-Fronting*

The Pearson's correlation allowed us to identify two sets of exceptions to the general spread of TH-Fronting in this community, and this section proposes an analysis to explain these exceptions. The sets are as follows:

(a) morphemes in which the dental in SSE precedes /r/ in a syllabic onset, e.g. *three* and *throw* (281 tokens in the corpus, 5% of all (th) sites). While such words do participate in TH-Fronting (in that variants such as [fɾi] do appear in the corpus), the set is exceptional because another variant is possible and is exclusive to this set. The variant is a palato-alveolar fricative [ʃ], so we consider the pre-onset /r/ context to be a different context of variation.

(b) the lexemes THING, THINK, and WITH, including derivatives of the first, such as *everything* and *anything* (4140 tokens in the corpus, 80% of all (th) sites).

With set (a), we propose that the assimilation of place, combined with the fact that /r/ is the only consonant that can follow the dental in an onset cluster, mark these words out as a special set. The palato-alveolar form is, to our knowledge, not found in non-Scottish varieties of English, suggesting some dialectal or sociolinguistic markedness, and the constraint in terms of onset-phonotactics suggests some

language-internal markedness. The Linguistic Atlas of Scotland (Mather and Speitel 1986) shows that similar variants, which are further reduced phonetically, such as a monomoraic onset with a devoiced liquid, are quite widespread in Scotland, but that variants with a palatal or palato-alveolar initial consonant are highly localised, with the only recorded instance being in Dunino, in east Fife[6]. However, we are also aware of such forms having appeared in a recent corpus of speech of young adults from Ayrshire (Pukli 2007) and in middle class speakers from Glasgow (Stuart-Smith, pc), so it may be that this form is more widespread than was previously recorded, or is becoming more frequent. Nonetheless, the present corpus suggests that for these Fife speakers at least, TH-Fronting may be sensitive to phonetic and phonological environment.

Some of the exceptions in set (b) are more complicated. The behaviour of WITH is unremarkable, since the existence of zero-variants of the final consonant of this preposition have a long history in many varieties of (British) English; however, the effect of the existence of local variants for THINK, THING and derivatives of the latter does require analysis. In this community, and indeed in many others across central Scotland (see e.g. Chirrey 1999; Robinson 2005; Stuart-Smith and Timmins 2006), THINK and THING have highly localised variants [hɪŋk] and [hɪŋ], i.e. where the first consonant is a glottal fricative. This pattern is different from widespread TH-Fronting because the change from [θ] to [h] involves a reduction (in the form of lenition), and could legitimately therefore be counted as a phonetically motivated sound change (unlike th-fronting which seems instead to be a case of lexical diffusion). However, while the verb *think* is high frequency (largely as a result of its appearance in the grammaticalized prefab *I think*), tokens such as *everything* and

*anything* are of much lower frequency – and yet both sets are undergoing reduction from [θ] to [h].

We propose that the lenition here may in part be a result of the lexicalization of the *–thing* compounds. The fusion involved in the univerbation process often involves phonological reduction or loss (cf. the lexicalization of OE *hlafweard* 'loaf guardian' > *lord*, or of *forecastle* to *fo'c'sle*). This is perhaps more systematic than some cases of lexicalization, but is clearly an instance of this type of change. Such lexicalization is also likely to lead to greater entrenchment of the item as a unit (i.e. it will be more likely to be accessed holistically than compositionally) which will make it even further resistant to TH-Fronting. Since we are here viewing lexicalization as a diachronic process, we hypothesise that this may be a change in progress, with variants [θ], [h], [ʔ] and zero, this last being yet a further instance of lenition. Although the age range of our informants does not allow us to provide definitive evidence of a change in progress, evidence for lexicalization, at least in the *-thing* compounds, can be found when we examine the ways in which these variants pattern with lexical frequency.

[insert figure 3 about here].

This graph shows the variants of (th) in each of the *–thing* compounds in the WFHPB corpus. The data are arranged according to the frequency of the lexical item with *something* as the most frequently occurring *–thing* compound in these data. The pattern in the graph seems to support the proposal that this is a phonological change because the more frequently occurring *–thing* compounds are displaying evidence of

more reduced variants of (th) e.g. [ʔ] and the 'zero' variant. This pattern is statistically significant, as we can see in the chart below.

[insert figure 4 about here].

The correlation coefficient here is -.938 which is highly significant; there is a highly significant correlation between more frequent *–thing* compounds and the occurrence of the most phonetically reduced 'zero' variant of the (th) variable. In other words, there is evidence to suggest that the*–thing* compounds are involved in a different phonological change from TH-fronting; these lexical items may be becoming more lexicalized. If this does constitute a change in progress, it is much slower than TH-Fronting, since the latter seems to have spread rapidly in the community over one or two generations.

When these very high frequency lexical exceptions (discussed under (a) and (b) above) and their lower frequency derivatives are removed from the analysis, the simple correlation coefficient of the Pearson's correlation is 0.171 and is not significant. This means that there is no large correlation between th-fronting and token frequency in the remainder of these data.

## 3. EMBEDDING FREQUENCY WITHIN A LARGER FRAMEWORK OF LANGUAGE USE

### *3.1 The relative importance of frequency*

Our analysis of variation has so far only examined lexical frequency in isolation (as is typical in the frequency literature) and only using a simple linear correlation. More sophisticated statistical techniques such as multiple regression can, however, spot smaller effects because while computing the effect of one independent variable, it can explicitly control for the effects of all other independent variables, therefore reducing

the statistical 'noise' that can interfere in a simple linear correlation. We therefore decided to investigate the relative importance of the role of frequency in comparison with other factors in the spread of this sound change. To do this, we included a factor group testing for lexical frequency into a variable rule (hereafter varbrul) analysis using the statistical software package Goldvarb X (Sankoff et al. (2005)). A varbrul analysis can be used to ascertain the effects of various independent factors influencing the distribution of a dependent variable by means of stepwise multiple regression. The 'linguistic' factor groups that were included in the analysis are provided in table 1 and the 'social' factor groups are in table 2. These factor groups are discussed in more detail below.

[Table 1 about here]

[Table 2 about here]

**Preceding and following phonological context:** these factor groups coded for the possible effects of phonological context as an influencing factor in motivating (th) variation as no other studies of (th) have considered the effect of phonological context on this variation. We began by coding phonological context in detail with each individual segment as a separate factor but a number of cells were left empty or had very low cell counts (see Guy 1988:129-132 on the problems of low cell counts) and so it was necessary to collapse some of these factors together. We have chosen to represent the factors in this factor group on the front/back dimension. The main difference in articulation between [θ] and [f] is the position of the tongue in the vocal tract – the tongue occupies a fronted position in the mouth when articulating the

dental fricative. We therefore hypothesised that if th-fronting is influenced by phonological context, the dental variant may be more likely to occur either immediately preceding or following other fronted articulations.

**Word boundary:** again, as this has not been considered in previous studies of th-fronting, we were interested to discover if the variation in (th) was perhaps sensitive to morphological information such as the occurrence of a word boundary either immediately preceding or following the variable.

**Preceding [f] in the word:** the motivation for including this factor group was to test for a priming effect. Is the labiodental variant more likely to occur if the word in question has a labiodental voiceless fricative somewhere else (preceding the variable)?

**Place of (th) in the word:** Stuart-Smith and Timmins (2006) investigated the role of the lexicon in th-fronting in Glasgow and found that the labiodental variant occurs more frequently in word final position than word initially or word medially. They do not analyse the THINK/THING set of words separately from the remainder of (th) sites and they attribute their finding to the high frequency of these words and the fact that [h] occurs word initially (*thing*) and word medially (*everything*) in these lexical items. We were interested to know if this effect remains once the THINK/THING set of words are removed from the analysis of (th).

**Place of (th) in the syllable:** there is a great deal of typological evidence (see Kiparsky 2004) that place and manner features are frequently neutralized in syllable codas**.** We therefore might expect to find that th-fronting occurs more frequently in coda position.

**Lexical category:** Stuart Smith and Timmins (2006) note in their discussion of th-fronting in Glasgow that several of the lexical items that seem to be resisting th-

fronting in their corpora are ordinals and proper names and so we were interested to test this factor on these data.

**Individual speaker:** we initially coded each individual speaker separately because of the discomfort we felt at grouping sets of speakers into categories. In the end this was unmanageable as a factor group because there were too many speakers who either showed no variation (and so produced a ***knock out*** in varbrul – it is impossible to include these speakers in an analysis of variation as the data is not variable) or had a small number of tokens of (th). This factor group was therefore not included in the final analysis.

**Speaker sex:** based on the results of previous studies of th-fronting, it would appear that speaker sex should be a significant factor influencing variation. However, we wanted to test this variable and find out whether speaker sex is equally important in this community.

**Community of practice/Friendship group membership:** the friendship groups presented here are based on cliques found in a UCI NET analysis of the social structure of this group[7] and the labels in table 2 were given by the members of the community. These groups are therefore not categories that we have imposed on the community; they represent how the speakers themselves view the social organisation of the pipe band.

**Age:** age is a continuous variable and so the position of boundaries between factors is a somewhat arbitrary decision. The factors represented here are the result of several attempts to find the best fit of the model to the data.

**Length of time in the band:** we were aware that some individuals felt a great deal of affiliation to the band and had been a part of this organisation since they were young children while others had joined more recently or had played with other bands at different times in their lives. We attempted to measure the strength of affiliation to the

band by quantifying the amount of time each individual had spent there as a percentage of their life.

**Area of residence:** when we asked the speakers in the corpus if they were aware of linguistic variation in the group, they mostly responded that they were. When we asked them why they thought these differences existed, they were often quite insistent that this was simply the result of dialectal variation:

Extract 1

LC: see aw the folk in the band, dae they aw talk the same?

Bobby: nuh

LC: how no?

Campbell: aye you've got different eh dialects like it's amazing how many dialects are in Fife alone eh

We decided to test their intuitions with this factor group. Although table 2 shows a number of different localities (2 of which are not in Fife), the large majority of the group live in and around West Fife (only 4 of the 54 speakers do not live in Fife) and so the group is actually fairly homogeneous both socially and geographically.

In order to achieve a valid varbrul analysis, the factor groups must be 'orthogonal' (Guy 1988:136) i.e. there must be minimal overlap between the factor groups. This can often be difficult to achieve, for example there is a great deal of overlap between the factor groups 'place of (th) in the syllable' and 'place of (th) in the word' as the first consonantal segment of any morpheme is by default also in the onset position of a syllable. Independence of social factor groups is perhaps even more difficult to achieve as there is more potential for overlap (see Bayley 2002: 131). In this case, almost all of the social factors interacted substantially, as one might expect, given the multidimensional nature of social characteristics that make up any

given individual. We attempted to tease apart the different factors influencing variation by running the analysis multiple times, testing different social and linguistic variants against each other until all possible combinations had been exhausted[8]. We then compared the results of each analysis using a likelihood ratio test to find which provided the best 'fit' and therefore the best indication of the likely factors influencing this variation.

Because variable rule analysis requires discrete variants of all variables, it was unfortunately necessary at this stage to convert the continuous measurement of lexical frequency adopted in the Pearson's correlations into discrete categories. Rather than create arbitrary cut points in the data or force category divisions in order that the number of tokens in each was approximately equal, we plotted the results for (th): [f] against lexical frequency in a scattergram and searched for natural 'bunches' in the data (see figure 5).

[Figure 5 about here]

4 natural categories emerged in the data and these were coded into the following factors: 'low frequency' (up to 20 instances), 'low-mid frequency' (21 to 33 instances), 'high-mid frequency' (43 to 48 instances) and 'high frequency' (107 to 137 instances). While these categories do not contain an equal number of tokens or types, they represent the frequency categories that naturally emerged from the data[9].

When the factor group 'lexical frequency' is included in the varbrul analysis, the results are as presented in table 3.

[Table 3 about here]

Table 3 is organised to show the factor groups in the order of their significance on the variation. The results of the varbrul analysis show that lexical frequency is a significant factor influencing this variation but it is the last significant factor group to remain in the analysis. In other words, of all the factors influencing variation in these data, lexical frequency has the weakest effect. We will now take some time to interpret these frequency results in light of the generalisations that have emerged from the literature on lexical frequency before discussing the other factors affecting th-fronting in this community.

*3.2 Interpreting frequency correlations*

Results of the multiple regression and the Pearson's correlation on the WFHPB data suggest that the sound change *TH-Fronting* may be displaying patterns associated both with Bybee's RE *and* CE. On the one hand, in lexical items where the only possible variation is between the voiceless labiodental fricative and the voiceless dental fricative (i.e. in the 15% of the corpus that are not of the type WITH/THINK/THING or THR_ ), there is a significant RE pattern – speakers seem to be adopting the innovation more readily in words with higher token frequency than lower token frequency as we can see from the results at the foot of table 3. Bybee's explanation for the reduction effect is that language production is a neuromotor activity and as neuromotor activities are repeated, their execution becomes more efficient; gestures are thereby reduced which, in language, leads to assimilation and reduction processes. As words with a higher token frequency are more exposed to this reduction, they change more rapidly. However, while the RE may be suitable for general processes of assimilation and reduction, it cannot explain the pattern found here because TH-Fronting is not a reduction; it is not a phonetically motivated sound

change[10]. It is a straightforward case of lexical diffusion or " the abrupt substitution of one phoneme for another in words that contain that phoneme" (Labov 1994: 524).

There is also evidence of a type of CE in this data but only in the lexical items WITH, THINK, THING (and derivatives of THINK and THING). The Pearson's correlation showed that these very high frequency lexical items (and their derivatives) appear very infrequently with the [f] variant. Bybee typically invokes the CE to explain the pattern of change often found in grammatical and analogical change. The explanation for this pattern is that high frequency words become more entrenched and resist change on the basis of more productive patterns in the language. Again, however, lexical frequency can only go some way to explaining the pattern of frequency effects found in these data. For instance, while it is possible to explain the resistance to TH-Fronting shown by the very high frequency lexical items THINK, THING and WITH, frequency alone cannot explain why lower frequency derivatives of these lexemes (e.g. *thingmie*) are also not participating in this change.

In order to explain the different frequency patterns found in the WFHPB data with the generalisations proposed by Phillips (2006), it is necessary to assume that the lexical items that allow TH-Fronting require no lexical analysis beyond phonetic coding in order for the change to take place. This means that there is no need for the speaker to access either more abstract schemas such as the item's word class, or phonotactic generalisations, in order to implement this change; the change simply requires 'shallow access' (Phillips 2006: 75) to the phonetic form of the word. However, if TH-Fronting was simply a change that affected the realisation of the word form and required no 'deeper' level of lexical analysis in its implementation then we would expect to find no correlation between TH-Fronting and word class. The analysis of variation presented in table 3 suggests that this is not the case. Not

only is there a statistically significant relationship between, for example, word class and the realisation of the labiodental fricative, which suggests that more abstract generalisations or schemas are important to the spread of this sound change, but this factor groups actually accounts for more of the variation in this data than lexical frequency (and so is perhaps even more important in the spread of TH-Fronting).

3.2 *Returning to the analysis of variation*

After lexical frequency, the factor group which displays the least effect on the variation in th-fronting while still remaining significant is lexical category. These results support the proposition made by Stuart-Smith & Timmins (2006) that ordinals and place names may be more resistant to the spread of TH-Fronting than other lexical items as ordinals, place names and proper names favour retention of the dental fricative. Kunter (2007) has also found that proper nouns and place names behave differently than other lexical items as they show significantly less variation in stress placement.

Syllable structure/place of (th) in the word is the next most important constraint on the variation in th-fronting. The results for this factor group suggest that when (th) occurs in syllable/word initial position, it favours the dental fricative and when it occurs syllable/word finally, the labiodental is more likely to occur. As stated, there is a great deal of typological evidence that marked place and manner features are frequently neutralized in favour of their unmarked values in syllable codas e.g. in the final devoicing of obstruents in German, most Slavic languages, Catalan, Turkish, Korean, and in many dialects of English. Following Steriade (2004), Kiparsky (2004) suggests that a plausible reason for coda neutralization might be the low perceptual salience of the relevant featural distinctions in the syllable coda. We therefore might expect to find that TH-Fronting occurs more frequently in coda

position. Stuart-Smith and Timmins (2006) found that in Glasgow, the labiodental variant occurs more frequently in word final position than word initially or word medially. They attribute this to the high frequency of the *think/thing* set of words and the fact that [h] occurs word initially (*thing*) and word medially (*everything*) in these lexical items and do not analyse the THINK/THING set separately as we have done. The results from the WFHPB corpus would appear to suggest, however, that this effect remains even after the *think/thing* set of words are removed from the analysis.

The next most significant constraint on variation in (th) is the factor group which codes for a priming effect at the phonological level[11]. As Hudson (2007: 37) explains, priming experiments in the psycholinguistics literature show that a preceding word or segment can prime a following word or segment by making it more quickly retrievable. This happens at all levels of the grammar: words with similar phonological, morphological, syntactic or semantic structure will act as primes. For instance, *verse* primes *nurse, hedges* primes *hedge* and *bread* primes *butter* (see Hudson 2007: 38-39 for a detailed discussion of this phenomenon and for further examples). In this case, if the lexical item in question contains a labiodental fricative somewhere before the variable (such as in the lexical item *fourth*) then the variable itself is much more likely to be realised as a labiodental fricative. This would suggest a type of priming effect. Hudson invokes this phenomenon as evidence for the existence of a network structure of organisation in the mind in which every node in cognition is ultimately connected to every other node. In this model, the activation of one node will automatically induce the secondary activation of an infinite number of connected nodes. One possible interpretation is that priming takes place through a process of *'spreading activation'* (Colins and Loftus 1975, Langacker 1987:385). Nodes in the network that the speaker perceives to be similar in some way are more

closely linked in cognition and the further apart two nodes are in the network; the less likely they are to activate each other. In this case, it is possible to invoke spreading activation to explain the fact that a preceding [f] in the lexical item seems to prime the realisation of the variable (th): [f]. When a particular instance of the [f] node is activated, it spreads activation to phonologically similar nodes making the variable (th) more likely to be realised as [f][12].

Finally, the factor group 'community of practice/friendship group membership' substantially outranks all other constraints on the variation. In other words, there is a very strong correlation between the use of the labiodental fricative and membership in a particular social group in this community[13].

The main advantage to employing a varbrul analysis of the variation in TH-Fronting is that we have the capability to model both social and linguistic factors simultaneously impacting on a speaker's choice of variants and to rank their relative strength and significance. We have seen from the above analysis that a number of linguistic and social factors are important in influencing this variation and that the role of lexical frequency is perhaps less important than it would appear from the number of studies which examine it in isolation. However, in order to interpret these findings, it has been important to embed the discussion of lexical frequency within a larger theoretical framework of language use. In other words, in order to fully understand the results that are apparent in the WFHPB data, it is necessary both to consider a wide range of possible motivating factors that may be influencing this variation and to interpret the results of these factors within an explanatory socio-cognitive framework such as a usage-based model (Kemmer and Barlow 2000).

## CONCLUSION

The patterns of TH-Fronting in WFHPB do show frequency effects but these are not entirely in line with what has been predicted based on previous research on the progression of sound change and lexical frequency. Clark (2008) finds a similar result to this in an examination of vocalic change in a sub-section of WFHPB. Furthermore, when lexical frequency is considered independently of all other influences on variation (as is typically the case in frequency research) the correlation between lexical frequency and TH-Fronting is different than when other more relevant contributory factors are also included in the analysis. These results together indicate that there is rarely ever one single motivating factor responsible for the spread of a linguistic change, a fact which we have been aware of since at least the 1960s:

> Explanations of language which are confined to one or other aspect – linguistic or social – no matter how well constructed, will fail to account for the rich body of regularities that can be observed in empirical studies of language behaviour (Weinreich et al. 1968:188).

We propose that it is vital not only to consider the role of lexical frequency as an explanatory factor in the spread of sound change but to embed this further within a wider theoretical framework of language use.

AUTHORS' CONTACT DETAILS

*Linguistics and English Language*

*School of Philosophy, Psychology and Language Sciences*

*University of Edinburgh*

*Dugald Stewart Building*

*3 Charles Street*

*Edinburgh EH8 9AD*

*Scotland, UK*

*L.Clark-8@sms.ed.ac.uk*

*Graeme.Trousdale@ed.ac.uk*

REFERENCES

Abramowicz, Lukasz. 2006. 'Sociolinguistics Meets Exemplar Theory: Frequency and Recency Effects in (ing)'. Paper presented at NWAV 35, Columbus, Ohio.

Baayen, R.H., Piepenbrock, R., & Gulikers, L. 1995. The CELEX Lexical Database (Release 2) [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [Distributor].

Bayley, Robert. 2002. 'The quantitative paradigm'. In J.K Chambers, Peter Trudgill and Natalie Schilling-Estes (eds.) *The handbook of language variation and change. Oxford,* 117-14. Blackwell Publishers,

Borgatti, Stephen P., Martin G. Everett and Linton C. Freeman (2002). Ucinet for

Windows: Software for social network analysis. Harvard: Analytic Technologies.

Bybee, Joan. 1985. *Morphology: A study of the relation between meaning and form.* Amsterdam: John Benjamins.

Bybee, Joan. 2000. 'The phonology of the lexicon: evidence from lexical diffusion'. In Suzanne Kemmer and Michael Barlow (eds.), *Usage-based models of language,* 65-86. California: CSLI Publications

Bybee, Joan. 2001. *Phonology and Language Use.* Cambridge: Cambridge University Press.

Bybee, Joan. 2002. 'Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change'. *Language Variation and Change* 14, 261-290.

Bybee, Joan. 2006. 'From usage to grammar: the mind's response to repetition'. *Language* 82 (4), 711-33.

Bybee, Joan. 2007. *Frequency of use and the organisation of language*. Oxford: Oxford University Press.

Chirrey, Deborah. 1999. 'Edinburgh: descriptive material' in Paul Foulkes and Gerard Docherty (eds.) *Urban Voices: accent studies in the British Isles*, 223-229. London: Arnold.

Clark, Lynn .2008. 'Re-examining vocalic variation in Scottish English: a Cognitive Grammar approach'. *Language Variation and Change* 20(3): 255-273.

Clark, Lynn .In prep. *Variation, change and the usage-based approach*. PhD thesis, University of Edinburgh.

Clark, Lynn and Graeme Trousdale .2008. 'Phonological variation in a Scottish community: Methods and theory in Cognitive Sociolinguistics'. Paper presented at the symposium *Approaches to Variation and Change in English*, Otto-Friedrich-Universität, Bamberg.

Collins, A.M. and E. Loftus. 1975. 'A spreading-activation theory of semantic processing'. *Psychological Review 82*: 407-28.

Dinkin, Aaron. 2007. 'The Real Effect of Word Frequency on Phonetic Variation'. *Penn Linguistic Colloquium* 31.

Eckert, Penelope. 2000. *Linguistic Variation as Social Practice: The Linguistic Construction of Identity at Belten High*. (Language in Society 27.) Oxford: Blackwell.

Foulkes, P. & Docherty, G.J. 2006. 'The social life of phonetics and phonology'. *Journal of Phonetics* 34(4): 409-38.

Guilquin, G. 2007. 'The cognitive reality of frequent verb-noun combinations: An empirical study'. Paper presented at the 10th International Cognitive Linguistics Conference (ICLC), Jagiellonian University, Krakow

Guy, Gregory R. 1988. 'Advanced varbrul analysis'. In Kathleen Ferrara, Becky Brown, Keith Walters and John Baugh (eds.) *Linguistic change and contact,* 124-36. Austin, TX: Department of Linguistics, University of Texas at Austin.

Guy, Gregory R. 1993. 'The quantitative analysis of linguistic variation'. In Dennis Preston (ed.) *American dialect research,* 223-249. Amsterdam: John Benjamins.

Hay, Jennifer. and Baayen, Harold. 2002. 'Parsing and productivity.' In Booij, G. E. and Van Marle, J. (eds), *Yearbook of Morphology 2001,* 203-235. Kluwer Academic Publishers, Dordrecht.

Hay, Jennifer. 2001 'Lexical Frequency in Morphology: Is Everything Relative?' *Linguistics* , 39 (6), 1041-1070

Hudson, R.A. 2007. *Language Networks: the new word grammar.* Oxford: Oxford University Press.

Kemmer, Suzanne and Michael Barlow. 2000. 'A usage based conception of language' in Suzanne Kemmer and Michael Barlow (eds.) *Usage-based models of language*. California: CSLI Publications

Kerswill, Paul. 2003 'Dialect levelling and geographical diffusion in British English'. In D Britain and J Cheshire (eds.) *Social Dialectology. In honour of Peter Trudgill,* 223-243. Amsterdam: Benjamins

Kiparsky, Paul. 2004. 'Universals constrain change, change results in typological generalizations'. Ms, Stanford University. Available at http://www.stanford.edu/~kiparsky/Papers/cornell.pdf

Kunter, Gero .2007 'Within-speaker and between-speaker variability in compound stress assignment'. Paper presented at the 2nd International Conference on the Linguistics of Contemporary English (ICLCE), University of Toulouse.

Labov, William. 1994. *Principles of linguistic change vol. 1: internal factors.* Oxford: Blackwell.

Labov, William. 2006. 'A sociolinguistic perspective on sociophonetic research'. *Journal of phonetics* 34, 500-515.

Labov, W., Ash, S., & Boberg, C. 2006. *Atlas of North American English: Phonetics, phonology and sound change.* Berlin: Mouton/de Gruyter

Langacker, R.W. 1987. *Foundations of Cognitive Grammar volume 1: theoretical prerequisites*. Stanford: University of Stanford Press.

Langacker, R.W .1988. 'A usage-based model'. In Brygida Rudzka-Ostyn (ed.) *Topics in Cognitive Linguistics (current Issues in Linguistic Theory 50).* 127-61. Amsterdam: Benjamins.

Langacker, R.W .2000. 'A dynamic usage-based model', in Suzanne Kemmer and Michael Barlow (eds.) *Usage-based models of language*. California: CSLI Publications

Macafee, Caroline. 1983. *Varieties of English around the World: Glasgow*. Amsterdam: Benjamins.

Mather, James Y. and Hans-Henning Speitel. 1986. *The Linguistic Atlas of Scotland, Volume III: phonology.* London: Croom Helm.

Mathews, Tanya. 2005. *Discourses of intergroup distinctiveness among adolescent girls.* Unpublished PhD dissertation: Cornell University.

Phillips, Betty. S. 1984. 'Word frequency and the actuation of sound change'. *Language* 60: 320-42.

Phillips, Betty. S. 1998. 'British vs. American –ate and –ator: convergence, divergence, and the lexicon'. *American Speech* 73: 160-77.

Phillips, Betty. S. 2006. *Word Frequency and Lexical Diffusion*. Basingstoke: Palgrave Macmillan.

Poplack, Shana. 1980. 'The notion of the plural in Puerto Rican Spanish: Competing constraints on (s) deletion'. InW. Labov (ed.), *Locating language in time and space*. New York: Academic Press. 55– 67.

Pukli, Monika. 2007. 'Vowels in Scottish English: durational patterns and the Scottish Vowel Length Rule'. Paper presented at the 2nd International Conference on the Linguistics of Contemporary English (ICLCE), University of Toulouse.

Robinson, Christine. 2005. 'Changes in the dialect of Livingston'. *Language and Literature* 14 (2), 181-193.

Sankoff, D; Tagliamonte, S.A & Smith, E. 2005. Goldvarb X. Available to download at http://individual.utoronto.ca/tagliamonte/Goldvarb/GV_index.htm.

Schmid, Hans-Jörg. 2007. 'Does frequency in text really instantiate entrenchment in the cognitive system?' Paper presented at the 10th International Cognitive Linguistics Conference (ICLC), Jagiellonian University, Krakow.

Steriade, Donca. 2004. 'The phonology of perceptibility effects: the P-map and its consequences for constraint organization'. In Kristin Hanson and Sharon Inkelas (eds.) *The nature of the word.* MIT Press: Cambridge.

Stuart-Smith, Jane and Tweedie, Fiona. 2000. *Accent Change in Glaswegian: A Sociophonetic Investigation* (Final report to the Leverhulme Trust). (F/179/AX).

Stuart-Smith, Jane and Timmins, Claire. 2006. 'The role of the lexicon in TH-Fronting in Glaswegian'. *Costerus. The Power of Words: Essays in Lexicography, Lexicology and Semantics.* No. 163, 171-184.

Tagliamonte, Sali. A. 2006. *Analysing sociolinguistic variation*. Cambridge: Cambridge University Press.

Travis, Catherine. E. 2007. 'Genre effects on subject expression in Spanish: priming in narrative and conversation. *Language Variation and Change* 19: 101-135.

Weinreich, Uriel, Labov, William and Herzog, Marvin. 1968. 'Empirical foundations for a theory of language change'. In Winfred P. Lehmann and Yakov Malkiel (eds.) *Directions for historical linguistics*, 95-188. Austin: University of Texas Press.

Wells, J.C (1982) *Accents of English: the British Isles*. Vol.2 Cambridge: Cambridge University Press.

Williams, Anne. & Kerswill, Paul. 1999. 'Dialect levelling: change and continuity in Milton Keynes, Reading and Hull'. In P. Foulkes & G. Docherty (eds.) *Urban Voices,* 141-162. London: Arnold,

FOOTNOTES

1. We are grateful to April McMahon, Rena Torres-Cacoullosto and Vsevolod Kapatsinski for their invaluable comments on an earlier draft of this paper. We would

also like to acknowledge the helpful comments of the audience at the 2nd ICLCE conference held in Toulouse where parts of this work were presented. Finally, we would of course like to thank the anonymous reviewers whose comments have helped us enormously.

2. We are aware that the literature on the relationship between lexical frequency and sound change is far greater than only these three sources but, given restrictions on space, we feel that these sources best represent a good mixture of the current research. Bybee (2007) is a collection of 15 single and co-authored papers that have appeared over the last 20 years or so on this topic and so it is a representative sample of her work on lexical frequency and phonological change; Phillips (2006) is a recent book on lexical frequency and phonology that also summarises a large amount of her early work and incorporates recent re-evaluations of this work; Labov (2006) is the only publication in which he deals with the topic of lexical frequency exclusively.

3. For further discussion of exemplars in phonological variation and change, see Foulkes and Docherty (2006).

4. This was modelled on a sorting task developed by Mathews (2005) in her research on the category labels that were given to adolescent girls in an American high school. The aim of the task was to understand how the informants grouped themselves and others in the community. The results of the sorting task were then taken as input data for a social network analysis of the community (using the software UCINET, Borgatti et al. 2002).

5. 'Awhing' is literally 'all-thing'; this is the Scots form of *everything*. It is a relatively uncommon variant (with only seven instances in the corpus, compared with 157 instances of *everything*) and in all cases in the corpus, it appears with the glottal fricative variant of (th) so we have chosen to transcribe it here as 'awhing'.

6. We are grateful to Keith Williamson for his advice on the distribution of these variants in the Linguistic Atlas of Scotland.

7. The details of this procedure are quite complicated and discussed in depth in Clark (In prep.).

8. For instance, on the first run of the multivariate analysis, we included the factor group CofP/Friendship group membership (but did not include the factor groups Age and Speaker Sex); on the next run we removed the CofP/Friendship group membership and included Age (but not Speaker Sex); on the next run we removed Age and included Speaker Sex, and so on.

9. See Clark and Trousdale (2008) for a discussion of the methodological problems associated with quantifying lexical frequency.

10. If this sound change was 'reductive' (i.e. if it was to be successfully attributed to a decrease in muscular activity of the tongue), we might expect to find some evidence of a correlation between the dental fricative and 'front' segments (such as front vowels) or the labiodental fricative and centralised or back segments (such as central or back vowels). Since there are no apparent correlations in the data (see section 4) an argument in favour of construing this sound change as 'reductive' is not tenable.

11. See, for example, Poplack (1980) and Travis (2007) for discussions of priming in studies of language variation and change.

12. We are grateful to Kapatsinski (pc.) for pointing out that activation need not necessarily *spread* from /f/ to prime (th):[f]. Activating /f/ early in the word may simply raise its activation level making its activation later in the word more likely (and so this would constitute a kind of identity or repetition priming).

13. The relationship between TH-Fronting and friendship group membership is discussed further in Clark (in prep).

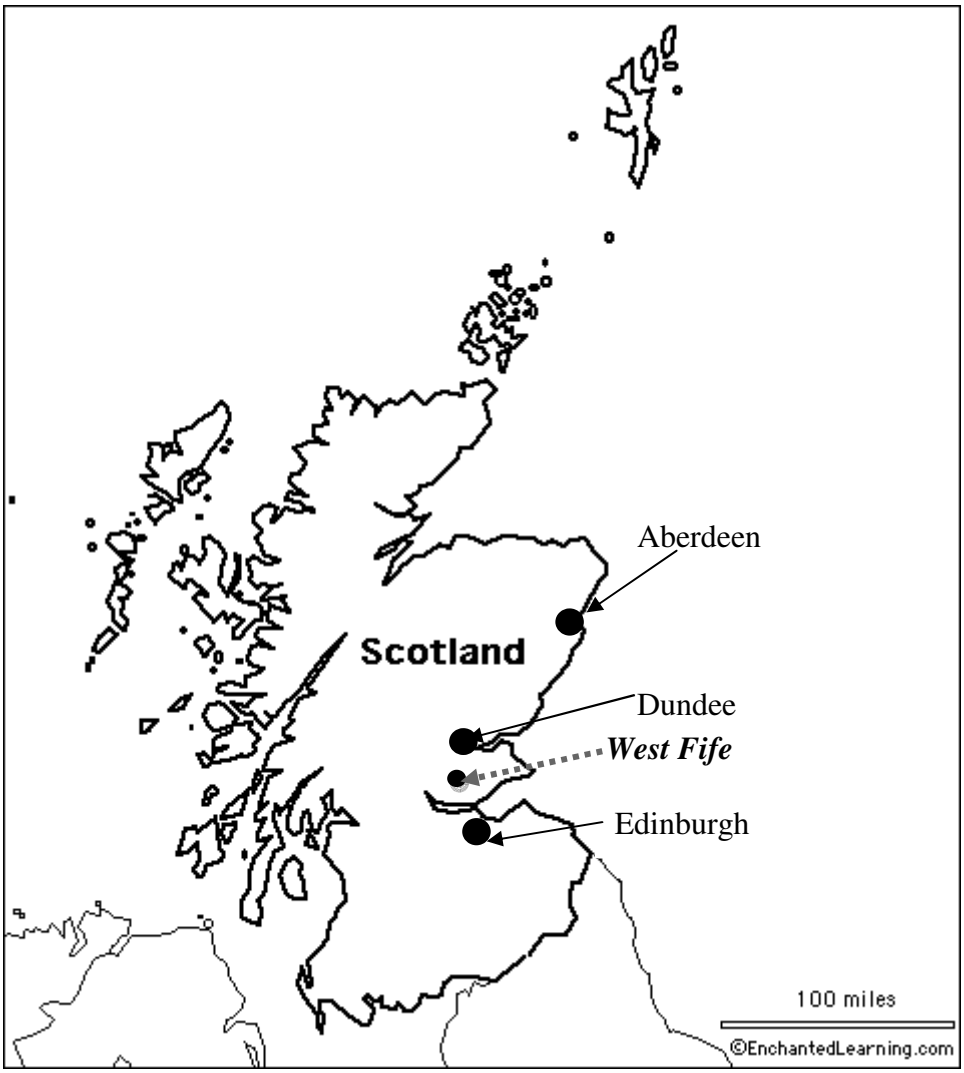


Figure 1
*Map showing the west fife area in relation to major cities along the east coast of Scotland*

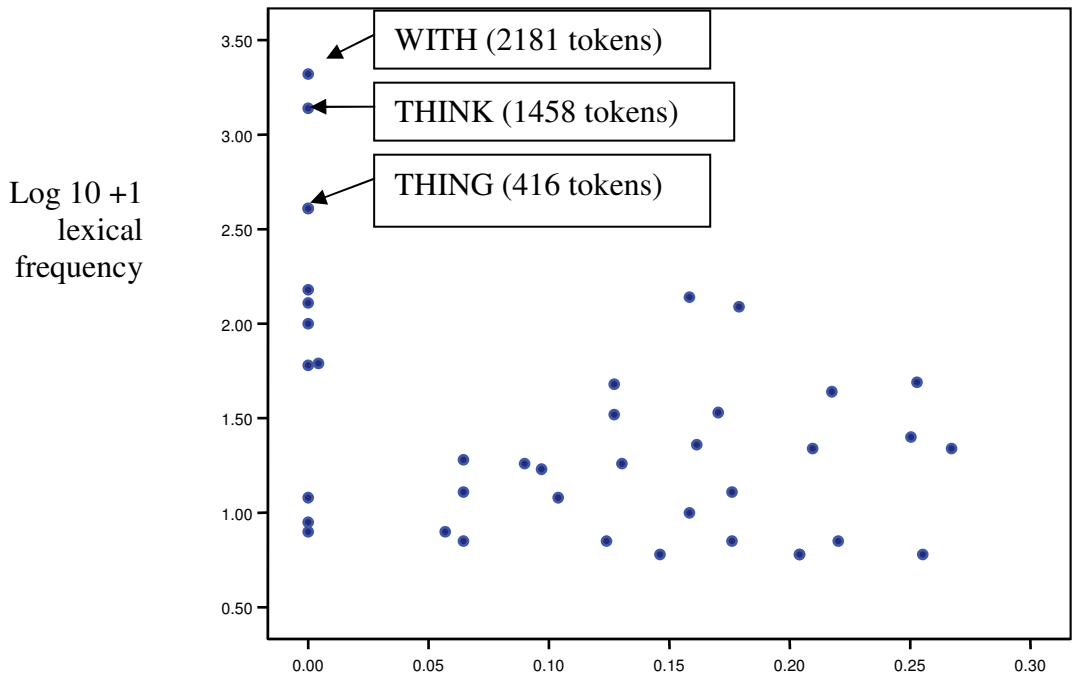


Figure 2
*Pearson's correlation of log lexical frequency and TH-Fronting.*

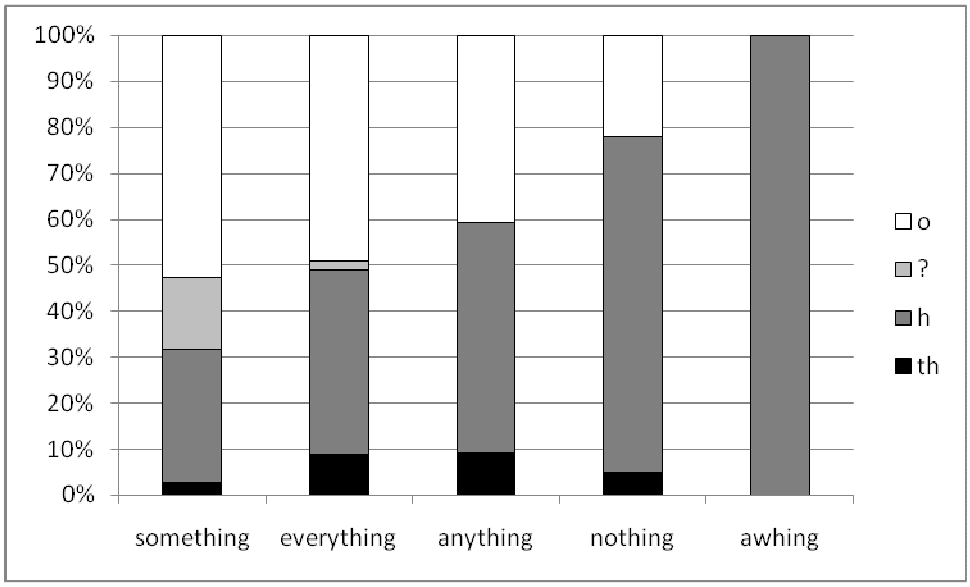


Figure 3
*Variants of (th) in the* –thing *compounds arranged by lexical frequency (from highest to lowest along the x axis)*[5]

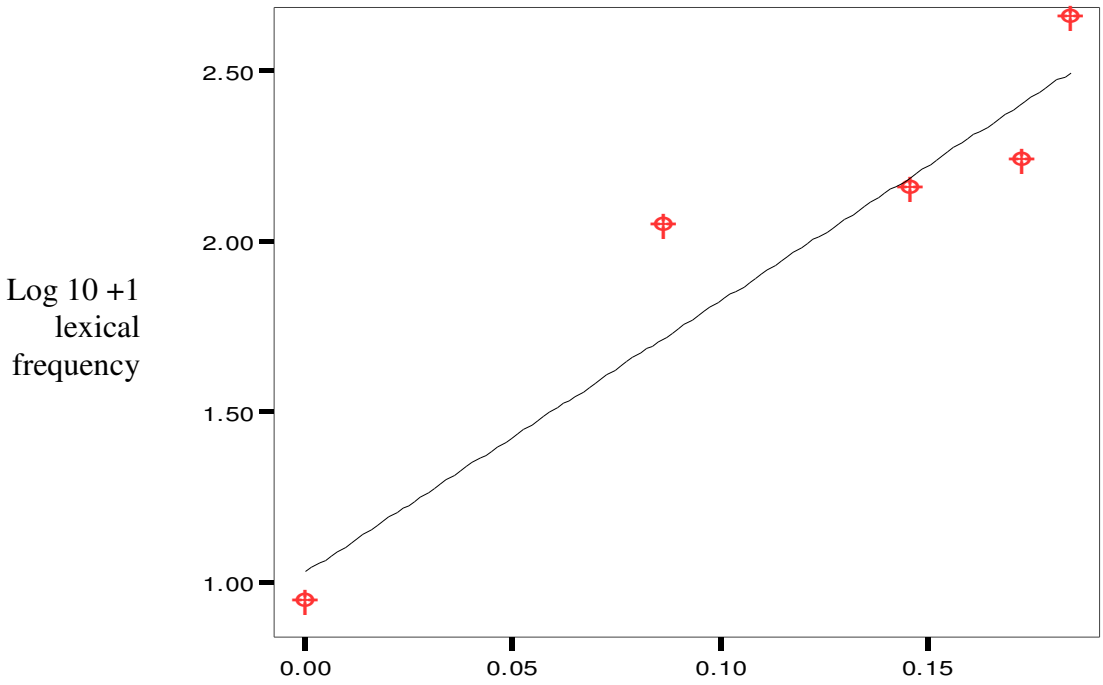


Figure 4
*Pearson's correlation of log lexical frequency and frequency of (th): 0 in the* –thing *compounds.*

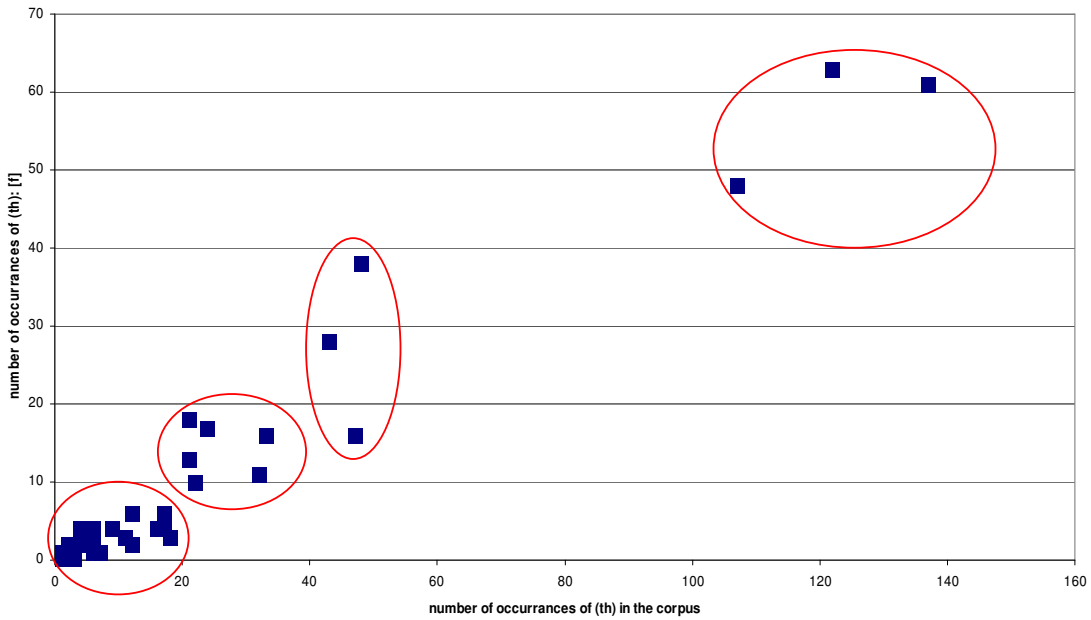


Figure 5
*(th): [f] plotted against lexical frequency in the WFHPB corpus (excluding lexical exceptions to TH-Fronting).*

| *Factor Group* | *Factors* | *Example Token* |
|---|---|---|
| **Preceding phon. segment** | Front vowel | *it's me an Billy an Kei**th** an that* |
| | Back vowel | *bad parts eh Glenro**th**es* |
| | Coronal consonant | *it wiz brilliant for aboot a mon**th*** |
| | Dorsal consonant | *we're aw on the same wavelen**gth*** |
| | Pause | *LC: so age-you're what fourteen?*<br>*A: **th**irteen* |
| **Following phon. segment** | Front vowel | *cos I'm **th**ick* |
| | Back vowel | *I've **th**ought eh everyhing else* |
| | Coronal consonant | *aboot **th**ree month never drinking* |
| | Dorsal consonant | *they've bo**th** got wives an children* |
| | Pause | *B: she just opens her moo**th***<br>*J: well I'll shut up then* |
| **Preceding word boundary** | Present | *mm hm, **th**irty year aulds* |
| | Absent | *we went tae see this mara**th**on eh* |
| **Following word boundary** | Present | *they've bo**th** got the same colour eh hair* |
| | Absent | *no Glenro**th**es* |
| **Preceding [ϕ]** | Present | *is it the four**th** wan ye need?* |
| | Absent | *eh Cowdenbea**th** pipe band* |
| **Place of (th) (syllable)** | Onset | ***th**ird* |
| | Coda | *same age bo**th** annoying* |
| **Place of (th) (word)** | Initial | *a **th**ought it wiz no bad* |
| | Medial | *what's it called-Me**th**il* |
| | Final | *Cos they're bo**th** in the same band* |
| **Lexical category** | Place names & proper names | *aye it's Kei**th*** |
| | Ordinals | *when he wiz in **th**ird an four**th** year* |
| | Other | *a **th**ought the jobs were starting* |

Table 1
*Linguistic factor groups for varbrul analysis of (th)*

| Corrected mean | 0.52 |
|---|---|
| Factor Group | Factors |
| Individual speaker | 54 individual factors, one for each speaker |
| Speaker sex | Male<br>Female |
| Friendship group membership | A “They act hard all the time”/ “fancy tune folk”<br>B “Tiny wee pipers”<br>C “The new folk”<br>D “Pipe band geeks”/ “Ex-Dream Valley”<br>E “comedians” /“Same dress sense, same music taste, same easy going attitude”<br>F “Fun/up for a laugh, not very serious”<br>G “that’s a fake ID son”<br>H “senior drummers”/”pipe band geeks”<br>I “one big happy family”<br>J “On the fringe”<br>K “13 goin on 30”<br>L “goths”/ “new lassie pipers”<br>M “Lazy PPl!”<br>N “Dollar lassies”<br>O “Under agers”<br>P “Novice tenor section ‘WILD’!!”<br>Q No CofP affiliation |
| Age | 12-15 years old<br>16-24 years old<br>25+ years old |
| Length of time in the band | < 10% of age<br>10-19% of age<br>20-29% of age<br>30-39% of age<br>40-49% of age<br>50+% of age |
| Area of residence | Lochgelly<br>Balingary<br>Lochore<br>Cardenden<br>Cowdenbeath<br>Falkland<br>Glenrothes<br>Scotlandwell<br>Rosyth<br>Dunfermline<br>Burntisland<br>Dollar<br>Leven<br>Dundee<br>Crossgates |

Table 2

*Social factor groups for varbrul analysis of (th)*

| | | | |
|---|---|---|---|
| Log Likelihood | | | -401.980 |
| Total N | | | 784 |
| | Factor weight | % of (th): [f] | N |
| **Community of practice/friendship group membership** | | | |
| A “They act hard all the time”/ “fancy tune folk” | 0.71 | 67 | 49 |
| B “Tiny wee pipers” | 0.95 | 93 | 56 |
| C “The new folk” | 0.89 | 85 | 59 |
| D “Pipe band geeks”/ “Ex-Dream Valley” | 0.10 | 7 | 27 |
| E “comedians” /“Same dress sense, same music taste, same easy going attitude” | 0.32 | 32 | 28 |
| F “Fun/up for a laugh, not very serious” | 0.75 | 75 | 24 |
| G “that’s a fake ID son” | 0.58 | 59 | 34 |
| H “senior drummers”/”pipe band geeks” | 0.09 | 9 | 76 |
| I “one big happy family” | 0.45 | 45 | 20 |
| J “On the fringe” | 0.21 | 23 | 57 |
| K “13 goin on 30” | 0.60 | 59 | 39 |
| L “goths”/ “new lassie pipers” | 0.51 | 55 | 87 |
| M “Lazy PPl!” | 0.31 | 30 | 78 |
| O “Under agers” | 0.48 | 44 | 32 |
| P “Novice tenor section ‘WILD’!!” | 0.79 | 78 | 45 |
| Q No CofP affiliation | 0.35 | 34 | 73 |
| Range | 86 | | |
| **Preceding [f] in the word** | | | |
| Preceding [f] | 0.81 | 68 | 22 |
| No preceding [f] | 0.49 | 48 | 762 |
| Range | 32 | | |
| **Syllable structure/place of (th) in the word** | | | |
| (th) in onset position | 0.37 | 38 | 486 |
| (th) in coda position | 0.58 | 55 | 298 |
| Range | 21 | | |
| **Type of lexical item** | | | |
| Place names and proper names | 0.42 | 48 | 351 |
| Ordinals | 0.42 | 39 | 324 |
| All other lexical items | 0.61 | 53 | 109 |
| Range | 19 | | |
| **Frequency of lexical item** | | | |
| Low frequency | 0.41 | 39 | 242 |
| Low-Mid frequency | 0.47 | 57 | 148 |
| High-Mid frequency | 0.53 | 60 | 139 |
| High frequency | 0.58 | 48 | 255 |
| Range | 17 | | |

Table 3
Multivariate analysis of the contribution of factors selected as significant to the probability of (th): [f]. Factor groups not selected as significant are not shown in this table.