

## Two directions of change in one corpus: Phonology vs morphosyntax in Tyneside English\*

Marie Møller Jensen, Aalborg University

**Abstract:** Cheshire et al. (2005) argue that different levels of language do not necessarily follow the same patterns of change over time. In an attempt to test this prediction, this article reports on a comparison between two quantitative corpus studies of Tyneside English which are partly based on the same data. The first study was carried out by Watt and investigated levelling in the phonological variables in the FACE and GOAT lexical set (Watt 2002) based on data collected in 1994. Watt found that speakers were abandoning broad local vernacular variants in favour of more regional or generally Northern forms. The second study was a study of seven morphosyntactic variables based on data collected in the 1960s, 1994 and 2007-2009. This study found that the variables under investigation were either stable over time or used more frequently in the most recent data. Thus, the comparison of the two studies shows support for Cheshire et al. (2005). This difference in the direction of change is explained by socio-psychological processes linking linguistic forms, the local Tyneside area and speaker identity. More specifically, it is argued that speakers imbue local vernacular variants with social meaning (Podesva 2006). This means that speakers create a social index (Silverstein 2003) which links linguistic forms and additional non-denotational meanings through processes of enregisterment (Agha 2003). This allows speakers to express an affiliation with their local area and all that it represents to them at a time when this area is undergoing vast changes.

**Keywords:** Enregisterment, identity, indexicality, morphosyntax, social meaning, Tyneside.

### 1. Introduction

Dialect levelling refers to the type of language change in which local dialects conform to regional varieties leading to the disappearance of more localized vernaculars (Trudgill 1986; Kerswill 2001, 2003). This change is often facilitated by the loss of rural dialect boundaries as people become increasingly mobile. Watt (2002) reported levelling in the phonology of Tyneside English based on data from the 1990s which is now part of the *DECTE* corpus (consisting of data from the 1960s, 1990s and now). Watt concluded that speakers were increasingly adopting supralocal variants as a means to retaining local membership whilst avoiding stigmatization through the use of broad vernacular forms. However, as Cheshire et al. (2005) remark, there is a lack of studies investigating levelling in other areas than phonology, and we should not expect levelling to affect different components of language in the same way. The study reported here aims to establish whether changes in the morphosyntax of Tyneside English can also be said to be levelling. It takes a quantitative approach and is based on data from the *DECTE* corpus. In short, the statistical tests performed map the frequencies of standard and vernacular forms of seven morphosyntactic constructions in order to establish any change over time. The results did not show that the variables were levelling. In fact, they either showed that the vernacular forms of the variables were used consistently over time or that their use was increasing. These results clearly do not follow the levelling hypothesis but support the observations of Cheshire et al. (2005) about the differences between phonology and morphosyntax.

The results of the morphosyntactic study are interpreted from a socio-cognitive perspective (e.g. Geeraerts 2005, 2010; Croft 2006, 2009). This approach to language and linguistic variation suggests that social and psychological factors can impact on speakers' language use. More specifically, it provides a framework for the interpretation of how changes in the urban landscape of

\* This research was carried out as part of my doctoral thesis which was fully funded by Northumbria University. I thank the editors of this issue of *Globe* and the anonymous reviewer for their comments which greatly improved the manuscript. Any shortcomings remain my own.

Newcastle upon Tyne (Miles 2005) may impact on inhabitants' affiliation with the area which may in turn affect their use of vernacular forms (Burbano-Elizondo 2008). It is thus hypothesised that the difference in direction of change is linked to social factors which, through a socio-cognitive link, affect speakers' language use. Jensen (2013) suggests that *salience* is the property which facilitates the attachment of social meaning to linguistic forms on a cognitive level. As this article is primarily concerned with the differences in the direction of change between Tyneside English phonology and morphosyntax found in the *DECTE* corpus, I will not go into the *salience* argument in any great detail here although I briefly return to it in the discussion of results.

This article will first discuss dialect levelling and Watt's study in more detail and provide definitions of key terms employed in the interpretation of the results of the statistical analyses. I will then introduce the study of morphosyntactic change in Tyneside English by firstly describing the corpus data used in the study, secondly, by defining the seven variables under study and, thirdly, explaining the method used to extract tokens and analyse their patterning. I then report on the results of the different statistical analyses before discussing them in light of both Watt's findings for Tyneside English phonology as well as the key terms defined in the second section. I conclude by a brief discussion of possible caveats and make suggestions for future studies.

## 2. Dialect levelling and linguistic identity

According to Watt & Milroy (1999: 31), phonological levelling has taken place in the Northeast “for at least forty years” and the broader Tyneside dialect can be seen as levelling towards a more regional standard called Tyneside English, a variety closer to the national standard. The fact that speakers want to appear modern but also still retain linguistic affiliations with their local area can be seen as a part of the argument for the claim made in Watt (2002) – namely that

(c)ontrary to claims that the distinctiveness of Tyneside English (TE) is eroding under the influence of a southern standard model, however, these generational differences are hypothesised to reflect TE's shift towards a northern, or north-eastern, regional standard. (Watt & Milroy 1999: 44-45)

However, as mentioned in the introduction, the stable or increased use of vernacular morphosyntactic forms is a change in the opposite direction of what is found in dialect levelling. Most studies within the framework of dialect levelling focus on changes in the phonology of localised dialects but the comparison presented here shows that studies investigating different levels of a variety using the same data might be a worthwhile undertaking. This furthermore supports Cheshire et al. (2005) who remark that there is a lack of studies investigating levelling in other areas than phonology and we should not expect levelling to affect different components of language in the same way.

### 2.1. Phonological levelling in Tyneside English

Watt's (2002) study of phonological levelling in Tyneside English was based on the *PVC* data which is now part of the *DECTE* corpus and described in more detail below. In his study, Watt compared the speech of 32 working class and middle class adults in two age groups: younger speakers (aged 18-25) and older speakers (aged 45+) and found that there was a significant difference in the vowels used by the two groups in words such as FACE and GOAT (see Table 1 below, from Watt 2002: 47):

Table 1: Phonetic variants of FACE and GOAT

	FACE	GOAT
<b>Type I (supralocal)</b>	e:	o:
<b>Type II (local)</b>	ɪə	ʊə
<b>Type III (national)</b>	eɪ	oʊ ə:

The supralocal variants are classed as 'generally northern' or Scottish pronunciations which Watt also calls “‘mainstream' northern variants” (Watt 2002: 47). The local variants, here called type II, are more local than the variants in type I and are classed as typically north-eastern characteristics. These characteristics are often present in the speech of Tyneside speakers even when other traits of the accent are avoided. The national or standard variants in type III are almost never found in Tyneside English. The fourth vowel in the GOAT set is a monophthong which has been a variant in Tyneside English for some time (Watt 2002: 47). At the end of his study, Watt concludes that the use of type II variants is diminishing and that the supralocal forms are used instead. However, he states, there is some evidence that the national forms are increasingly being used by some members of the middle class (Watt 2002: 57). In his discussion of his results, Watt hypothesises that this use of levelled, regional forms can be seen as the middle ground between an old-fashioned, backward and stigmatized variety (broad Tyneside) and the national standard which allow the speakers to retain local membership as well as allowing them a part in the modern and globalized nation at the same time (Watt 2002: 57-58).

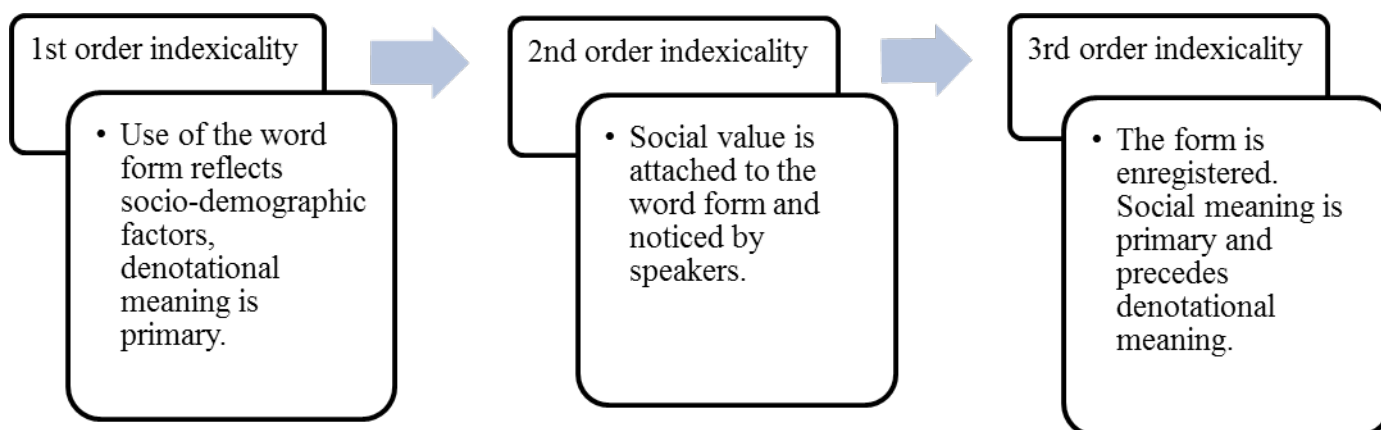
As I have already mentioned above, the results of the morphosyntactic study revealed markedly different results. While Watt found the broad Tyneside diphthongs to be used less by the younger generation of speakers, the morphosyntactic study found that local vernacular morphosyntactic forms, such as *divn't* and *hoy*, were either used at a similar frequency over time or even increasing in use among younger speakers in the most recent data. The difference in direction of change between the morphosyntax and the phonology of Tyneside English raises a number of issues: one concerns the difference between phonology and morphosyntax on a cognitive level. How are these two levels structured and what are the differences in speakers access to or awareness of these levels? This is discussed further in section 5. Another important issue which should be addressed is that of the reason for the increase in use of local morphosyntactic forms. The levelling hypothesis provides an explanation for the loss of broad local phonological variants, but it obviously cannot be applied to the morphosyntax. I suggest that this change is linked to changes in the urban environment of Newcastle upon Tyne which has undergone a transformation from industrial heavyweight to a haven of leisure facilities (Miles 2005). I argue that the difference in the direction of change found in the *DECTE* corpus for phonology and morphosyntax can be linked to a socially meaning-bearing function of the local morphosyntactic forms (Silverstein 2003, Podesva 2006). Thus, this urban and cultural regeneration of the townscape seems to go hand-in-hand with the linguistic regeneration exemplified by the increased use of Tyneside vernacular morphosyntactic forms. In the subsection below, I will introduce the key terms involved in this argument. I return to and elaborate on this argument in more detail when I discuss the results of the statistical analyses in section 5.

## 2.2 Sociolinguistic concepts involved in shaping a linguistic identity

In this section, I introduce the key terms which I will use to interpret not only the results of the morphosyntactic corpus study but also in my discussion of the difference between the direction of change in Tyneside English phonology and morphosyntax. In my discussion, I will focus on the role of *social meaning* and the related terms of *indexicality* and *enregisterment*. The figure below shows how the different terms are related. It is structured around Silverstein's orders of indexicality, which

are explained further below, and the role of social meaning is elaborated on in the boxes in the front. We see that Agha's enregisterment is here synonymous with 3<sup>rd</sup> order indexicality. This is a simplification, but fits the purposes of this paper for the time-being. Enregisterment is also further explained below.

Figure 1: Indexical order and social meaning



*Social meaning* is the meaning which linguistic features accumulate over time in addition to their denotational meaning and can be defined as the enregistered value of forms which index a range of social ideals. Social meaning indexes different social characteristics and values which reflect different perceptions of the social world. These perceptions differ from person to person and thus the social meanings indexed by different linguistic forms are likely to differ between speakers. As a concept, social meaning is often used synchronically in linguistics as a means of describing what variation means to speakers in their daily lives and how social perceptions manifest themselves in language use. Thus, linguistic choices may be influenced by the social meaning of forms and the identities speakers wish to portray. In this way, social meaning may over time (through the medium of speaker choice) lead to diachronic change.

Linked to the notion of the social meaning of forms is the process of *enregisterment* which accounts for the diachronic process of the creation and accumulation of the social meaning of forms (Agha 2003). Enregisterment is based on the linking of forms and established social constructs in a speech community – something which has been described in great detail by Silverstein (2003) as *indexicality*. Both of these concepts prove very productive in the interpretation of the difference in direction of change which this article is concerned with.

Silverstein (2003) argues that in order to investigate the ways in which speakers relate linguistic features to socio-cultural values, and thus create social identities in interaction, we need to consider the concept of *indexical order*. In short, indexical order is the formulation of the observation that “*n*-th order indexical tokens” (i.e. linguistic features) have “contextual entailments” (i.e. social meaning) which are a consequence of “the ideological engagement users manifest in respect of the *n*-th order indexical meaningfulness” (the social values expressed and maintained by speakers) (Silverstein 2003: 193-194). Silverstein builds on this and adds the notion of “*n* + 1<sup>st</sup> order indexical value” which he defines as a competing structure of value which can be characterised as a distinct but overlapping form which directly indexes the ideological value (or social meaning) in communication (Silverstein 2003: 194). This “dialectic competition” between the two forms ultimately plays a role in linguistic change as the *n* + 1<sup>st</sup> (second) order indexical value replaces the *n*-th (first) order indexical value (Silverstein 2003: 194). Eckert (2008) builds on Silverstein's

indexical order and argues for the interpretation of the meanings of variables based on an *indexical field*. She states that a more fluid conceptualisation of the potential meanings of variables is needed as variation is inherent in the social meaning of variables (variants mean different things to different people in different situations). The indexical field is thus a

constellation of ideologically related meanings, any one of which can be activated in the situated use of the variable ... and each new activation has the potential to change the field by building on ideological connections. Thus, variation constitutes an indexical system that embeds ideology in language and that is in turn part and parcel of the construction of ideology. (Eckert 2008: 454)

As mentioned above, connected to the notion of indexical order and the social indexicality of forms is enregisterment which describes “processes through which a linguistic repertoire becomes differentiable within a language as a socially recognized register of forms” (Agha 2003: 231). Indeed, it can be argued that the  $(n + 1) + 1^{\text{st}}$  (or third) order indexical value of a linguistic form expresses the enregistered meaning of the form (which is shown above in Figure 1). Enregisterment is discussed in depth by Agha (2003) in connection with the emergence and spread of RP in Britain, a process he treats in much detail. However, the overarching theme of the article is how cultural values are socially produced, maintained and transformed through discursive interaction and how cultural value as a dynamic property applies to language, which is here seen as any other cultural form. Agha comments that when non-linguists discuss accents they are not actually talking about specific sound patterns but rather about “a system of contrastive social personae stereotypically linked to contrasts of sound” (Agha 2003: 241-242). As an example of this, he mentions that RP is “enregistered in cultural awareness as part of a system of stratified speech levels linked to an ideology of speaker rank” (Agha 2003: 242). He proposes that the transmission of cultural values across a population takes place through discourse, through what he calls a *speech chain*. This he defines as

a historical series of speech events linked together by the permutation of individuals across speech-act roles in the following way: the receiver of the message in the  $(n)^{\text{th}}$  speech event is the sender of the message in the  $(n+1)^{\text{th}}$  speech event, i.e. where the terms 'sender' and 'receiver' ... are variable names of interactional roles, specified in different ways at different points along the speech chain. (Agha 2003: 247).

A similar argument could be made for Tyneside English (and indeed other non-standard vernaculars which are thriving in today's Britain) where local word forms have been linked to an ideology. In this case not an ideology of speaker rank but rather an ideology of localness. Thus, social meaning (or cultural value, in Agha's terminology) is constructed through language but also part of the language itself.

Johnstone (2009) presents an empirical study of Pittsburgh English and clearly links the use of local linguistic forms with speakers' expression of local identity. In this paper, she focuses in particular on T-shirts featuring words or expressions believed to be written representations of Pittsburghese. Johnstone argues that the consumption of these T-shirts is part of a process which has generated the idea that a distinct Pittsburgh dialect exists. These T-shirts not only put the dialect on display, they also infuse the local vernacular with (social) value and create a standardized form of the vernacular. Finally, they also create a link between local vernacular speech and particular social meaning (Johnstone 2009: 157). The Pittsburghese T-shirts rely on enregisterment to find their market. Johnstone notes that it is only individuals who are able to recognise Pittsburgh speech as

distinct from other varieties and who link it with “authentic local identity” (Johnstone 2009: 168) who will find the shirts funny or appealing. In addition, the print on the shirts has to be recognised as a representation of Pittsburghese and thus rely on already enregistered forms. However, third order indexicality (or enregisterment) of a range of linguistic forms is also a product of these shirts through the display of forms, infusion of value, creation of a standard, and link with social meaning.

What we see from the above descriptions of indexicality and enregisterment is that they are useful theoretical terms to consider in the discussion of language and social identity. What they make clear is that language exists not of itself but shapes and is shaped by speakers' social identity. Speakers are seen as active participants in the construal of social meaning through their language use and it is precisely this link between the social and the cognitive aspects of language which the socio-cognitive approach to language captures. The social, then, is not just an afterthought but very much part and parcel of what is conveyed by speech. Foulkes & Docherty (2006: 419), writing in the area of sociophonetics, summarise this in the following way: “Indeed, the interweaving of sociophonetic and linguistic information in speech is so complete that no natural human utterance can offer linguistic information without simultaneously indexing one or more social factor”. In their 2006 paper, Foulkes and Docherty explore the area of sociophonetic variation, drawing on findings from some of their own previous studies on Tyneside English, among other varieties. They also discuss sociophonetic variation from the perspective of first language acquisition, again focusing studies of data collected in Newcastle. They suggest an exemplar-based model in their account of how social and linguistic information may be acquired, stored long-term and accessed in on-line processes of production and perception although they also make clear that it is not clear, at present, how sociophonetic information is represented cognitively and how it is processes in comparison with other types of information.

They present insights from studies on variation on the segmental, suprasegmental, and subsegmental level and also present evidence (from Newcastle and Derby) supporting the ability of phonetic contrast to index social information. In other words, phonetic variation across speakers is not merely be a reflection of physiological differences between males and females but is meaning-bearing and can be perceived by listeners. The study looked at preaspiration and voicing in both Newcastle and Derby and found that while, in Newcastle, extended voicing was used more often by males than females (across class and age) and preaspiration was used mostly by young females (across both working and middle class). On the other hand, preaspiration was not found at all in Derby and extended voicing showed no significant social effects.

### 3. Data, variables and method

The aim of the morphosyntactic corpus study was to investigate the frequency of use over time for seven morphosyntactic variables: sentential negation with *do*, first and second person pronouns, sentential negation with *can*, and the verbs *go*, *throw*, and *told*. The variables are introduced in more detail below. Essentially, this study employs standard variationist methodology by considering linguistic variables and their envelopes of variation, i.e. how many different ways there are of saying the same thing (Labov 1972: 323). This methodology is based on Labov's work in the 1960s, which set the precedent for investigating the patterning of variation in language. This article adopts standard variationist annotation by placing the variable name in parentheses, e.g. (throw). Every linguistic variable has several variants, i.e. different ways that this variable can be expressed. In the example of (throw), the possible variants are *throw* (which is the Standard English form) and *hoy* (which is the Tyneside English form). Standard variationist annotation sees variants given in italics.

The seven variables included in this study were selected based on two different types of sources: firstly, scholarly literature (Beal 1993, 2004, 2010; Beal et. al. 2012) on the Tyneside dialect was consulted and, secondly, further examples were found in what can be described as *popular dialect literature*. This type of literature describes a genre of books and pamphlets, most of

which were published in the 1970s, which are either wholly or partly written in what is claimed to be *Geordie* or Tyneside English. These books often deal with aspects of the dialect (e.g. in the book *Larn yersel' Geordie* by Scott Dobson) or Geordie culture (e.g. *Scott Dobson's Geordie Recitations, Songs and Party Pieces*) in a humorous fashion and are aimed at visitors to Tyneside (as would be the case for *Larn yersel' Geordie*) as well as Tyneside speakers (perhaps particularly expatriate Geordies). However, the choice of variables was also limited by methodological considerations involving the types of search possible using the software programme R to search through raw (i.e. not annotated or parsed) corpus data. As for the clearly lexical variables, the criterion was that the lexical forms had to be particular to Tyneside. For the morphosyntactic variables, the criterion was that the variables displayed non-standard morphosyntactic forms in the syntactic environments under study.

Due to time constraints, this study does not consider any constraints of the variation (neither internal, external nor extra-linguistic) although it recognises that further investigations into these issues would yield valuable results. Before progressing to the study proper, however, it is important to make clear here how morphosyntax is understood as there can be an overlap between what constitutes morphosyntactic variation and different forms of lexical items in non-standard varieties. According to Crystal (2009: 315), morphosyntactic forms are “grammatical categories or properties for whose definition criteria of morphology and syntax both apply, as in describing the characteristics of words”. An example of this is the singular/plural distinction of nouns. The grammatical number of a given noun affects the corresponding verb when the noun is in the subject position, i.e. number affects syntax. In addition, if a noun is in the plural, it takes a plural ending (e.g. *-s*), i.e. number also affects morphology. In this way, variation in morphosyntactic variables affects both the surface forms (i.e. addition of plural *-s* on nouns) as well as the underlying syntax (i.e. the requirement for subject-verb concord where a singular noun requires a singular verb). The grey area between lexicon and morphosyntax arises as it is sometimes difficult to establish whether a variable is an example of one or the other. Lexical forms will most likely have less impact on the underlying syntax (although there are clearly reasons for why a speaker chooses one lexical form over another) than a morphosyntactic variable, which is why definition and classification is important. Although the main focus is on frequency change in standard and vernacular morphosyntactic forms, a few clear lexical variables have been included in the corpus study (e.g. (throw) which has the TE form *hoy*). However, some of the variables investigated here also fall into the grey area between morphosyntax and lexicon (an example is the variable (*go*) which has the TE form *gan*).

Each variable is described in more detail in section 3.2 below and the origins of the vernacular forms given. It is likely possible to argue that some variants of a variable reflect a clear synonymous relationship whereas others may display simple lexical form variation due to their etymology (and others again are examples of morphosyntactic variation). Considerations of this kind, while valid and insightful, not only raise issues outside the scope of this paper (differentiation between morphology and lexicon as briefly described above, the constitution of a synonymous relationship versus simple variation in form) but they are perhaps also less relevant in a study of this kind for two reasons. First and foremost, this study is interested in binary pairs of standard and non-standard forms regardless of whether they can be classed as synonyms or not and whether they are strictly morphosyntactic or more towards the lexical domain. Secondly, what is of the essence is thus the vernacular quality of the variants which ultimately is a quality wholly determined by the Tyneside English speakers (i.e. a form is only a vernacular form if it is perceived to be one and thus indexes locality to some extent). This means that the status of the variants as morphosyntactic, lexical, synonym or form variant becomes less important.

### 3.1. Data

The data used for this corpus study is the *Diachronic Electronic Corpus of Tyneside English* (*DECTE*, Corrigan et al. 2010-2012) which is comprised of three subcorpora: *The Tyneside Linguistic Survey* (*TLS*), the *Phonological Variation and Change* corpus (*PVC*) and the *Newcastle Electronic Corpus of Tyneside English 2* (*NECTE2*). The data stored in these three subcorpora were collected in the 1960s (*TLS*), in 1994 (*PVC*) and in 2007-present (*NECTE2*, the data included in this study was collected 2007-2009). The data stored in these corpora is interview data. The table below outlines the earliest and latest possible birthdates for the speakers in each corpus (adapted from Barnfield 2009). While this study does not consider informant age or year of birth in the analysis of change and variation, this table has been included here to give the reader an impression of just how many years the data manages to capture. The *DECTE* corpus is truly a unique resource in that it incorporates local speech data from informants born in the late 1800s until now.

Table 2: Overview of data

Corpus and years collected	Younger speaker birthdates (age 17-34)	Older speaker birthdates (age 35+)
<b>TLS 1965-1970</b>	1935- 1968	1895- 1934
<b>PVC 1991-1994</b>	1954- 1977	1911- 1953
<b>NECTE2 2007-2009</b>	1967- 1990	1923- 1966

Before proceeding to the introduction of the individual subcorpora, it should be highlighted that the data stored in these corpora is not perfectly matched. A few ways in which the data differ include geographic spread (the *TLS* data is exclusively from Gateshead, the *PVC* data is exclusively from Newcastle, and the *NECTE2* data is from a larger area which can be described as Tyneside); age range (although this has been normalised for this study, i.e. informants have been separated into similar age groups across the three subcorpora); operationalization of social class (this is often a tricky subject in sociolinguistic studies, see also Jensen (2013) for a discussion of the issues of social class in general in the North of England); the number of speakers in each social cell (e.g. no old MC speakers in *NECTE2*, only 1 old male WC speaker in *PVC*); fieldwork methods and protocols of transcription.

#### 3.1.1. *The Tyneside Linguistic Survey*

The data in this corpus was collected in the late 1960s in Gateshead, which is on the southern bank of the river Tyne. The data-driven approach pioneered in the survey is still employed today and is recognised for its empirical benefits to hypotheses of language variation and change (Corrigan et al. 2000-2005). A large amount of work has been put into restoring and securing the *TLS* data, some of which had been lost and some badly damaged. Today, 37 files, which contain complete interviews with informants and full transcriptions, are available and all were used in this study. The data files also provide social information about each speaker (age, gender and detailed social class based on level of education) and, on the basis of this information, the speakers were separated into the following categories:

Table 3: Overview of the *TLS* data

WC				MC				Total
Young (17-34)		Old (35+)		Young (17-34)		Old (35+)		
Male	Female	Male	Female	Male	Female	Male	Female	
3	5	5	6	5	6	4	3	37



The interviews consist of a guided conversation between an interviewer and one informant, averaging 30 minutes in length, some interviews taking on a more relaxed conversational style and others a more formal question – answer format (Corrigan et al. 2000-2005).

### 3.1.2. *The Phonological Variation and Change in Contemporary Spoken English* corpus

This data was collected in Newcastle on the northern bank of the river Tyne between 1991 and 1994. The methodology used was broadly similar to that commonly employed in variationist sociolinguistic fieldwork today which means that it differs from that employed by the *TLS* fieldworkers. The interviews last around 60 minutes and involve informal conversations between a pair of friends or relatives. The *PVC* corpus consists of a total of 18 files each featuring 2 speakers and all were included in the study. The social distribution is shown below in Table 4. This data was also used in Watt's (2002) study of phonological change which is discussed further below.

Table 4: Overview of the *PVC* data

WC				MC				Total
Young (17-34)		Old (35+)		Young (17-34)		Old (35+)		
Male	Female	Male	Female	Male	Female	Male	Female	
5	5	1	3	6	4	7	5	36

### 3.1.3. *The Newcastle Electronic Corpus of Tyneside English 2*

The material in the *NECTE2* corpus is collected by undergraduate and postgraduate students at Newcastle University and it consists of several data files, each containing an interview between an interviewer and two speakers (using the same methodology as the *PVC* corpus), a word list, and a reading passage. The style of the interviews is informal with minimal participation of the fieldworker and the speakers are, for the most part, closely acquainted. The interviews last around one hour. The files selected for this study were collected in 2007, 2008 and 2009 and the speakers were from either Newcastle or Gateshead in order to ensure maximum comparability with the speakers in the *PVC* and *TLS* corpora. A total of 24 files (48 speakers) were selected and the social distribution of speakers is given below in Table 5:

Table 5: Overview of the *NECTE2* data

WC				MC				Total
Young (17-34)		Old (35+)		Young (17-34)		Old (35+)		
Male	Female	Male	Female	Male	Female	Male	Female	
8	6	6	7	9	6	0	6	48

## 3.2. *Variables*

The seven variables included in the corpus study are described in more detail below and the variants included in the study listed.

### 3.2.1. (do + NEG)

The Tyneside English contracted form for this construction is *divn't* (also represented as *divvnt*) and, according to Beal (1993: 192), the auxiliary *div* (for *do*) is unique to Tyneside. Beal further states that the auxiliary *div* can occur in both positive and negative present tense statements and tag questions and that the phonological form *div* is never used for the main verb *do* (see sentence 1 below). Rowe (2007:361) adds that the positive form of the auxiliary *div* is rarely used except by conservative speakers and speakers using certain linguistic features as in-group markers

(particularly a group widely identified across the Tyneside region as *charvas*<sup>1</sup>). Finally, *divn't* does not occur in the third person singular which is always *doesn't*, according to Beal (2004: 124) (although Rowe (2007: 365) gives the form *dizn't*). Whilst there is clear evidence that *divn't* is the dominant vernacular form of (do + NEG) in the Tyneside area, other non-standard forms can be found as well (see e.g. Cheshire et al. (1993) and Buchstaller & Corrigan (2011)). The variants included in this study were: *do*, *don't*, *don-t*, *div*, *divn't*, *divn-t*, *divn*, *does*, *doesn't*, *doesn-t*, *dinna*, *divven't*. The examples below are taken from the corpus:

- (1) what div I like to do in my spare time well... (tls28, male, old, WC)
- (2) and that you know and this pott singer I divn't care for that fellow I like to hear it sometime but as for watching it on television I don't care much for that you know (tls14, male, old, WC)
- (3) I don't know how I've got this... I divn't knaa where all my money's gone (necte2 07-08/N/ML/159, male, young, MC)

### 3.2.2 Pronouns

Tyneside English is by no means alone in displaying variation in the pronoun system. In fact, this is a common occurrence in regional varieties of English (Trudgill & Chambers 1991: 7; Beal 2010: 39). This study only deals with the first and second person personal pronouns, although TE pronouns differ from those of Standard English in a number of ways. Some of these differences are also found in other regional dialects (such as using the object pronoun in the subject position in compound subjects, using *which* with a personal antecedent) and some are particular to Tyneside English (such as adding *-self/selves* to the vernacular possessive forms of pronouns throughout the paradigm giving forms such as *meself* and *theirselves* (Beal 1993: 205-207, 2004: 117-119).

- (First person pronoun): In Tyneside English, we find that the standard paradigm has been completely reorganised apart from the first person subject, as can be seen from the table below (Beal 1993: 205):<sup>2</sup>

Table 6: First person pronouns in Standard and Tyneside English

	Standard	Tyneside
<b>Subject singular/plural</b>	I / we	I / us
<b>Object singular/plural</b>	Me / us	Us / we
<b>Possessive singular/plural</b>	My / our	Me / wor

Beal (2010:42-43) discusses pronoun exchange in regional varieties of English and defines it as follows: “[p]ronoun exchange’ is the term used to refer to a phenomenon whereby what would, in Standard English, be the subject form is used in the object form and vice versa” and continues to note that in the Northeast, only the first person plural forms have been exchanged. However, as can be seen from Table 6 above (which is based on Beal 1993) the

1 A term used in Newcastle to denote groups of “tough” young people most often from a lower socioeconomic background known for their use of distinctive linguistic features (to signify group membership) as well as particular dress-code (branded sports apparel). The term has been absorbed into general English in recent years (it was *Word of the Year* in 2004) and now denotes members of the ‘underclass’ across Britain although the distinctive dress-code of sports apparel and tendency to cause havoc in town centres is maintained (Rowe 2007, Hayward and Yar 2006).

2 It should be noted here that in the most recent publication about North-eastern English, Beal et.al. (2012: 52) report Tyneside English to have the form *we* in the plural subject form, i.e. the same form as Standard English, however, the data for the corpus study reported here was compiled and analysed prior to this resource becoming available.

object singular and possessive forms are different in Tyneside English. Beal (2010) further comments that *we* (pronounced with a weakened schwa vowel) in the object position is more frequent than *us* used in the subject position. The variants included in this study were: *we*, *us*, *me*, *my*, *our*, *wor*, *mi*. While the first person singular subject form *I* formed part of the initial data collection, it was excluded from the analyses as it is the same form in both Standard and Tyneside English and accounted for more than half of the initial 40,000+ tokens which were collected. Below are a few examples taken from the corpus data:

- (4) Keeps us on my toes (necte2, 07-08/G/DM/456, young, male, MC)
- (5) And he used to buy we like alcohol and that (necte2, 07-08/G/LR/195, young, female, WC)
- (6) and they constantly had me mam ganning up to the school to talk about us and stuff (necte2, 07-08/N/PS/243, young, male, WC)
- (7) Oh yeah, we're great friends with wor next door neighbours (necte2, 07-08/N/VL/3892, old, female, MC)

- (Second person pronoun): The vernacular form of the second person personal pronoun is *yous* (in both singular and plural, see below) in Tyneside English. This form has most likely been introduced by Irish immigrants and the form is also found in other northern urban varieties, e.g. Liverpool and Manchester (Beal 2010: 40-41). An older vernacular form in TE is the singular subject form *ye* (plural form: *yees*) which is thought to be a remnant from the Early Modern English period. For speakers who have the *ye* form, the second person pronoun paradigm has distinct forms for all four positions (where Standard English has *you* in all four environments). However, the *ye* and *yees* forms were very rare in the corpus data (Beal 1993: 205, 2004: 118, 2010: 40). As this study is strictly interested in the change in frequencies of non-standard forms over time, coding did not differentiate between the different vernacular forms used<sup>3</sup>. As we can see from the table below, there is an overlap in forms between Tyneside English and Standard English in the singular object position. Both Englishes have *you* in this position which makes it impossible to determine whether it is the vernacular or standard pronoun which is being used. In the coding of data, all occurrences of *you* were labelled as Standard English. Whereas this holds the potential to be misleading due to the ambiguous data, time constraints and the somewhat 'raw' format of the data meant that this seemed the best solution to this issue as opposed to leaving out tokens in the singular object position.

Table 7: Second person pronouns in Standard and Tyneside English

	Standard	Tyneside
<b>Subject singular/plural</b>	You / you	(Ye) yous / yous
<b>Object singular/plural</b>	You / you	You / yous (yees)

The following variants were included in this study: *you*, *yous*, *ye*, *yees*, *ya*. The examples

<sup>3</sup> I acknowledge that this coding scheme hides internal patterns of variation across the different syntactic environments and social categories, however, as mentioned previously, this study is purely interested in changes in frequencies of vernacular forms over time. Furthermore, the corpus data had only a handful of tokens of the forms *ye* and *yees*, although this could be due, in part to the differences in transcriptions across the three corpora.

below are taken from the 1990s data:

- (8) it's just yous were good weren't you oh apart from that time yous collapsed (pvc09a, male, young, MC)
- (9) I know my mam says “yous are stupid yous are letting her manipulate you again making you feel guilty when you shouldn't have to feel guilty” (pvc12a, female, young, MC)

### 3.2.3. Verbs

This final category contains the following four verbal variables: *can* + negation (which is *canna*), the vernacular form *gan* for Standard English *go*, TE *hoy* for Standard English *throw*, and finally the form *telt* for Standard English *told*. The criteria for the selection of the four variables in this category were that they had to be either lexical forms particular to Tyneside (as is the case for *hoy*) or display non-standard morphosyntax (as is the case for *canna*). As mentioned previously, *gan*, but also *telt*, occupy the grey area between morphosyntax and lexicon.

- (can +NEG): According to Beal (1993: 199, 2004: 123), speakers of Tyneside English tend to opt for uncontracted constructions of in sentential negation with the auxiliaries *have*, *be*, *will*, and *can*. The TE form for Standard English *cannot* is *canna* (also reproduced as *cannae*). The negative particle *na* or *nae* is also found extensively in Scotland (Trudgill & Chambers 1991: 49; Dictionary of the Scots Language 2005). The variants included in this study were: *can not*, *cannot*, *can't*, *canna*, *cannae*, *can-nae*, *can-not*, *can-na*, *canne*, *can-ne*. The examples below are from the corpus:
 

(10) Yeah that's how different we are I would prefer going on holiday even though I can-nae sit in the sun 'cause I burn loads. (necte2 08-09/N/SG/456, young, female, WC)

(11) aye I'm sick of telling them if somebody else can hear it as well as you it canna be doing you no good (pvc18b, old, female, WC)

(12) I've just always quite liked it here I cannae think of a down side (necte2 Tessa.Durby, old, female, MC)
- (go): According to Beal (1993: 192), the Tyneside English form *gan* is a “lexically distinct verb” which is not found in Standard English. It is attested in the Survey of English Dialects (Upton et al 1994) in the imperative and in exclamations such as *gan to hell*, *gan on*, and *gan off* from Durham, York and Norhumberland. According to Oxford English Dictionary, *gan* stems from the Old English infinitive (Oxford English Dictionary Online, “go, v.”) whereas Standard English has taken the Old Norse form. The table below shows the present tense paradigm for *gan* based on the occurrences in the three corpora:

Table 8: (go) in Standard and Tyneside English

	Standard	Tyneside
1 <sup>st</sup> person singular	I go	I gan / gans
2 <sup>nd</sup> person singular	You go	You gan
3 <sup>rd</sup> person singular	He / she / it goes	He / she / it gans
1 <sup>st</sup> person plural	We go	We gan / gans

<b>2<sup>nd</sup> person plural</b>	You go	(no occurrences)
<b>3<sup>rd</sup> person plural</b>	They go	They gan / gans

As we can see, there is some variability in the endings in the first person singular and first and third person plural. According to Beal (2010: 32), some Northern varieties of English have *-s* throughout the present tense paradigm (and not just in the third person singular as is the case for Standard English). However, the matter is complicated somewhat by the 'Northern Subject Rule' which states that "the verb takes *-s* in the plural where the subject is a noun or noun phrase, but not when it is a pronoun adjacent to the verb" (Beal 2010: 32). Based on the data used for this study, it seems that the two rules are in competition and that Tyneside speakers differ in which forms they prefer when. The following variants were included in the study: *go*, *goes*, *goin*, *going*, *gan*, *gans*, *gannin*, *ganning*. The examples below are from the corpus:

- (13) aye we used to play in the street you ca you couldn't gan anywhere else to play (tls06, old, female, WC)
- (14) we often gan on about it now (tls03, old, female, WC)
- (15) drink bottles when I gan in there (pvc01b, young, male, MC)
- (16) the insurance gans down ((doon)) a tenner every week? (necte2 07-08/N/PM/85, young, male, WC)
- (17) Ah that music was ganning till half two last night did you hear it? (necte2 07-08/N/ML/159, young, male, MC)
- (throw): The Tyneside verb for 'throw' is *hoy*. It is relatively infrequent, however, it is a verb which is often mentioned as a 'stereotypical' Geordie word, e.g. in the oft-quoted phrase "Hoy the hammer over here" (e.g. see BBC, 2008). It is attested in Wright (1898) as a verb found in Northumberland, Durham and Cumbria meaning "to throw" with the first entry dated 1969. A similar entry is found in the Survey of English Dialects (Upton et al 1994). In Wright (1898), *hoy* is also mentioned as an exclamation occurring in other, more southern parts of England (Devon, Kent, Nottinghamshire, Leicestershire and Lancashire). Furthermore, it is also attested in the Oxford English Dictionary and the definition given can be linked to the Tyneside English use for 'throw' albeit tentatively. The OED lists *hoy* with the meaning "[t]o urge on or incite with cries of 'hoy!'; to drive or convoy with shouts" (Oxford English Dictionary Online, "hoy, v.") and gives examples from as far back as 1536 and includes an example by Robert Burns, the famous Scottish poet, from 1786:
- (18) They *hoy't* out Will, wi' sair advice.

Based on the data used in this study, *hoy* seems to follow the regular verb paradigm as can be seen from the table below (based on the corpora used in this study):

Table 9: (throw) in Standard and Tyneside English

	<b>Standard</b>	<b>Tyneside</b>
<b>1<sup>st</sup> person singular</b>	I throw	I hoy

<b>2<sup>nd</sup> person singular</b>	You throw	(no occurrences)
<b>3<sup>rd</sup> person singular</b>	He / she / it throws	(no occurrences)
<b>1<sup>st</sup> person plural</b>	We throw	(no occurrences)
<b>2<sup>nd</sup> person plural</b>	You throw	(no occurrences)
<b>3<sup>rd</sup> person plural</b>	They throw	They hoy

Other forms which occurred in the data were *hoying* as well as *hoyed* (used as past participle in the construction *got hoyed* and in the past tense *he hoyed it*). The variants included in the study were: *throw*, *throws*, *threw*, *thrown*, *throwing*, *throwin*, *hoy*, *hoys*, *hoyed*, *hoying*, *hoyin*. The examples below are both from the corpus data:

- (19) that's it you used to hoy a few currants in (pvc02a, old, male, MC)
- (20) even when there was lasses in my college I never got put with any of them I got hoyed straight in with the lads (pvc06a, young, male WC)
- (21) and the other lass was a bit thin because eh you have to hoy the boxes though you see (tls37, old, female, WC)
- (22) Oh he got hoyed out didn't he, aye! (necte2 07-08/G/JF/123, young, male, MC)
- (told): The final variable in this category is the past tense form of the verb *tell* where Tyneside English has the regular suffix *-t* (which gives the form *telt*) rather than following the irregular paradigm of Standard English which has *told* (Beal 2010: 31). As this study is purely concerned with mapping frequencies of use over time of Standard and vernacular forms, it does not distinguish between past tense and participle forms (Tyneside English has *telt* in both constructions and Standard English has *told*). The variants included in this study were: *telt*, *told*. The examples below show how the vernacular form was used by speakers in the corpus:
 

(23) but you telt me it was a fact (pvc06b, young, male, WC)

(24) it was him who telt me (tls28, old, male, WC)

(25) he telt us he was having a party but he didn't tell us like when (pvc01a, young, male, MC)

### 3.3. Method

Tokens from the corpora were extracted using the program R (R Development Core Team, 2011), the coding of the tokens was done manually in Microsoft Excel 2010 and statistical analyses were carried out using SPSS 19.0.

The corpus data was structured so that each line began with a speaker code and the full turn of the informant followed and each line ended with either the speaker code again or a code signalling the end of a turn. If the speaker turn ran over more than one line, it was divided into two (or more) lines at a natural point, all beginning and ending with the speaker code (or an end of turn code). The three subcorpora were merged to form one large corpus of approximately 700,000 words which was used as the basis for the token collection. It was possible to identify which corpus each token was from on the basis of the speaker code. The tokens were extracted by R in the following way: the

corpus file was narrowed down to only those lines that contained speech of informants by using the *grep()* function. The corpus was then further narrowed down to only those lines which contain matches with the search terms again by using *grep()*. The function *gregexpr()* was used to get a complete list of all matches (as some lines contained more than one match) and the lines with matches were then split into three (see below) by using the functions *rep()*, *sapply()*, *unlist()*, and *substr()* and the output was saved to a .txt file which was opened as a tab delimited table in Microsoft Excel 2010 where all further coding was done (for more information on the R code used, see Gries 2009: 138-140). The Excel table consisted of three columns: the first column featured the preceding context (from the beginning of the sentence including the speaker code up until the token), the second column contained the actual token (called 'match') and the third column the subsequent context (the remaining part of the line).

Figure 2: R output in Excel

1	PRECEDING_CONTEXT	MATCH	SUBSEQUENT_CONTEXT	INFORMANT
2	<07-08/ N/ 11>	you	are a cunt Hobbes. <@> ... cheers mate ... † ... (N (3))	<07-08/ N/ 11>
3	<07-08/ N/ 11>	yees	wore blazers? ... where did you go to school at? <07-	<07-08/ N/ 11>
4	<07-08/ N/ 11>	you	you were really bad <@> <07-08/ N/ 11>	<07-08/ N/ 11>
5	<07-08/ N/ 11>	you	would be getting that letter ... you would be getting	<07-08/ N/ 11>
6	<07-08/ N/ 11>	you	did not bother with The Who? <07-08/ N/ 11>	<07-08/ N/ 11>
7	<07-08/ N/ 11>	you	are kidding? <07-08/ N/ 11>	<07-08/ N/ 11>
8	pre .. just going to type it up ... bang .. done .. so	you	just get a print-out now and you are done ... aye ... I	<07-08/ N/ 11>
9	ing .. done .. so you just get a print-out now and	you	are done ... aye ... I would never .. prepare for it thou	<07-08/ N/ 11>
10	/ 11> (continues American Accent) .. soccer .. do	you	want a game of soccer? ... (reverts back to standard	<07-08/ N/ 11>
11	f soccer? ... (reverts back to standard accent) do	you	fancy a game of football? aye .. aye. <@> <07-08/ N/	<07-08/ N/ 11>
12	they play that .. they have got that umm .. have	you	seen that .. Dos Santos .. kid {.. the Brazilian guy ... <	<07-08/ N/ 11>
13	<07-08/ N/ 11> .. like a .. five goal deficit .. then	you	have to write the other guy a letter of apology .. to s	<07-08/ N/ 11>
14	etter of apology .. to say <@> .. saying how poor	you	are at Pro Evolution Soccer. <07-08/ N/ 11>	<07-08/ N/ 11>
15	is .. and concepts ... it seems .. in your third year	you	have got to read like .. scientific journals like .. all th	<07-08/ N/ 11>
16	ead like .. scientific journals like .. all the time ..	you	don't get any text books. ... (N (3)) <07-08/ N/ 11>	<07-08/ N/ 11>

If a line contained more than one match, each of these were stored in separate lines. All searches for the linguistic variants specified that these forms had to occur between word boundaries. This meant that forms such as *you're* were also included but occurrences of variants within other words (e.g. in *yourself*) were ignored. The use of word boundaries is also why, when searching for the variants of (do + NEG), the different negated forms had to be specified as a search for *do* alone with word boundaries would not return instances of *don't* and *doesn't* and a search for *do* without specifying that it should occur within word boundaries would return a multiple of other lexical items (such as *doing*, *down*, *donation*, *bulldog*). The same R code was used for the extraction of all the tokens for all the variables with only the search terms being different. All variables were kept separate throughout and thus the search was carried out once for each individual variable.

As the number of tokens collected for the different variables varied greatly, different statistical methods were used to investigate frequency changes across the three corpora. All tests, however, were concerned with mapping the frequency differences between the three groups, *TLS*, *PVC* and *NECTE2*. Not only must the tests be able to establish whether the patterning of tokens changes across the three groups, they must also be able to tell us whether the differences are statistically significant and between which of the groups the differences are largest. The two categories *sentential negation* and *pronouns* were analysed using parametric tests (ANOVA) and the variables in the final category, *verbs*, were analysed using non-parametric tests (chi-squared and Kruskal-Wallis). These are described in more detail below. Due to the use of different tests, the data needed to be prepared differently following the initial extraction from the corpora.

### 3.3.1. Parametric tests

Parametric tests (ANOVA) were used for the analyses of variables (do+NEG), (first person

pronoun) and (second person pronoun). As the number of tokens collected for each speaker varied greatly for each of these variables and because a very large number of tokens were collected overall, a random sample of 10 tokens (for *do* + NEG) or 20 tokens (for the pronouns) per speaker was selected and coded for source corpus and whether the token was standard or Tyneside English. Based on this selection, each speaker was given a *vernacular score* (for (first) and (second person pronouns) this score was between 0-20 and for (*do* + NEG) between 0-10) which simply comprised of the number of vernacular tokens in the random selection for each speaker. A between-groups (or independent) analysis of variance (ANOVA) was then carried out on the basis of the vernacular score. For the (*do* + NEG) variable, the initial search returned around 3,400 tokens. 10 tokens were then randomly selected for each speaker using Excel's RAND function to ensure a balanced and equally representative sample. Out of the 120 speakers in the corpus, 17 speakers produced less than 10 instances of sentential negation with *do* and were left out of the final sample. This left 103 informants (*NECTE2*=43, *PVC*=30, *TLS*=30) and a total of 1030 tokens.

For the first person pronoun, R initially returned over 40,000 tokens. However, this included the singular nominative form *I* which made up more than half of the total number of collected tokens. As this form is the same in both standard and Tyneside English, all tokens of *I* were removed from the data set. 20 tokens were then randomly selected from each speaker; however, 7 speakers had produced less than 20 instances of the first person pronoun so these were left out of the final sample. This left 113 speakers (*NECTE2*=45, *PVC*=36, *TLS*=32) and a total of 2,260 tokens. All selected tokens were coded manually according to variety (standard or vernacular) and grammatical role and number. This was necessary in order to determine whether the token is standard or vernacular due to the overlap in the pronoun paradigm. No instances of right-dislocated pronouns (e.g. *I don't like it me*) were included.

The total number of extracted tokens for the second person pronoun was a little greater than 15,000. It was not possible to eliminate any tokens from this data set because even though *you* does feature as both a standard and vernacular form it is also the only form in the standard. According to the paradigm, the only overlapping form between the standard and vernacular is the singular object (which is *you* in both varieties) and thus the only form which should be removed from the study if the method and line of argumentation used for the first person pronoun were to be replicated. However, in order to exclude all instances of the singular object form, all the tokens would have to be coded for number and position before the tokens could be removed. This was simply not very time efficient and thus all tokens were kept as the basis for the following random selection of tokens. Again, 20 tokens were selected from each speaker using RAND. Out of the 120 speakers in the collected corpus, 2 speakers were represented by less than 20 tokens in the data set and thus left out. This left 118 informants (*NECTE2*=47, *PVC*=36, *TLS*=35) and a total of 2,360 tokens. All instances of *you* were coded as 'standard'.

The table below shows the number of tokens included from the different corpora in the analysis of the first three variables:

Table 10: Distribution of selected tokens for ANOVA across corpora and variables

Corpus → ↓Variables	TLS		PVC		NECTE2	
	Standard	Vernacular	Standard	Vernacular	Standard	Vernacular
<b>(do + NEG) N=1030</b>	262	38	278	22	386	44
<b>(1<sup>st</sup> pers) N=2260</b>	591	49	651	69	785	115
<b>(2<sup>nd</sup> pers) N=2360</b>	697	3	713	7	865	75

### 3.3.2. Non-parametric tests

The rest of the variables (*canna*, *gan*, *hoy* and *telt*) were analysed using two different non-



parametric tests. Non-parametric tests were chosen as these can be used on smaller datasets as they do not rely on normally distributed data and do not make assumptions about the underlying population (Pallant 2007: 210).

The first test was the chi-squared test which tests for significant differences between groups of speakers over time. There is an issue, though, with applying chi-squared tests to a population of utterances (and not a population of speakers) where some speakers are represented by more tokens than others. This is because one of the (albeit few) assumptions for non-parametric tests is that all observations must be independent, i.e. each person may only be counted once (Pallant 2007: 211). However, it can be argued that for each token, the speaker had a choice between a vernacular and a standard form and thus each token represents a separate and independent speech act. This also means that what the chi-squared test reveals in this instance is variation across tokens rather than variation across speakers. As the number of tokens for the individual variables in the *verb* category was quite low, all tokens were included for all variables. The tokens were coded for the corpus they occurred in as well as whether the token could be classed as a standard or vernacular form and chi-squared tests were then carried out on the basis of this.

Because of the possible issue with the chi-squared test, Kruskal-Wallis tests were also carried out on the four variables. In short, Kruskal-Wallis is the non-parametric version of an ANOVA test (which was used to test the differences between the pronouns and (do + NEG)). The Kruskal-Wallis tests were based on a proportional score for each speaker which captured the proportion of vernacular tokens out of the total number of tokens collected for that speaker. The distribution of standard and vernacular tokens across the three corpora is given in Table 11 below:

Table 11: Distribution of tokens for non-parametric tests across corpora and variables

Corpus → ↓ Variables	TLS		PVC		NECTE2	
	Standard	Vernacular	Standard	Vernacular	Standard	Vernacular
<b>(can + NEG) N=260</b>	64	0	81	1	81	33
<b>(go) N=4567</b>	639	84	2146	93	1473	132
<b>(throw) N=86</b>	10	8	30	7	23	8
<b>(told) N=188</b>	28	2	78	13	62	5

#### 4. Results

I will present the results of the statistical tests (ANOVA, chi-square, and Kruskal-Wallis) in two separate sections. The first will detail the results of the ANOVA tests for the variables (do + NEG), (first person pronoun), and (second person pronoun) and the second section, 4.2, will detail the results for the four variables in the group *verbs* which were analysed using the two non-parametric tests, chi-squared test for independence and Kruskal-Wallis.

##### 4.1. Results of parametric tests

The table below summarises the descriptive statistics for the ANOVA test of the three variables discussed in this section. The ANOVA test compares the variability in scores between the three corpora (which is taken to be due to the time of collection) with the variability within each group (which is taken to be due to chance). Chance dictates that there will always be variation within the groups so for a significant result (expressed by a large  $F$  ratio), the variation between the groups must be larger than the variation within the groups as this would indicate that the independent variable which is being tested (here time of collection) is the cause of these differences (Pallant 2007: 242)<sup>4</sup>. As the ANOVA test is based on the vernacular values for each speaker, the means and

4 It should be noted here that this explanation accounts for the logic underlying the ANOVA test. It should be kept in mind that the variation within the groups may be structured and that the differences between the corpora may be

Standard Deviations also refer to these values. Interestingly, the maximum value reported for the first and second person pronouns in the *TLS* and *PVC* corpora are all very low, especially for the second person pronoun (1 and 2, respectively). The *Ns* given are the number of speakers included from each corpus.

Table 12: Summary of descriptive statistics for ANOVA

Corpus ↓ Variables →	(do+NEG)	(1 <sup>st</sup> pers. pronoun)	(2 <sup>nd</sup> pers. pronoun)	
<b>TLS</b>	N	30	32	35
	mean	1.27	1.53	0.09
	SD	2.449	1.796	0.284
	max	10	7	1
<b>PVC</b>	N	30	36	36
	mean	0.73	1.92	0.19
	SD	1.413	2.116	0.467
	max	5	6	2
<b>NECTE2</b>	N	43	45	47
	mean	1.02	2.56	1.60
	SD	1.858	2.981	3.221
	max	6	13	17
<b>FULL CORPUS</b>	N	103	113	118
	mean	1.01	2.06	0.72
	SD	1.933	2.443	2.164
	max	10	13	17

For (do + NEG), the one-way between-groups ANOVA did not show any effect of time of collection (or corpus) on the frequency of use of vernacular forms ( $F(2,100) = 0.568$ ,  $p = 0.568$ , N.S., effect size (partial eta squared) = 0.011 (no effect)). A non-significant result was also obtained for the second variable, (first person pronoun), ( $F(2,110) = 1.761$ ,  $p = 1.77$ , N.S., effect size (partial eta squared) = 0.031 (no effect)). The third and final variable in this section, (second person pronoun), is the first variable to show a significant effect. The one-way between-groups ANOVA indicated an effect of group (or time of collection) on speakers' use of vernacular forms ( $F(2,115) = 7.082$ ,  $p = 0.001$ , effect size (partial eta squared) = 0.110 (small effect size)). Post-hoc tests (Tukey HSD) indicate a significant difference between *NECTE2* and *PVC* ( $p = 0.007$ ) and between *NECTE2* and *TLS* ( $p = 0.004$ ). There was no significant difference between *PVC* and *TLS* ( $p = 0.973$ ).

These results show that there is no change in the frequency of use of (do + NEG) and (first person pronoun), however, there is an increase in use by speakers of the third variable (you), with speakers in the most recent data (*NECTE2*) using the vernacular form of the pronoun, *vous*, more frequently than in the 1990s (*PVC* data) and 1960s (*TLS* data). The results are summarised alongside those for the other four variables in Table 15 below on page 62.

#### 4.2. Results of the non-parametric tests

The four dependent variables in this category were (can + NEG), (go), (throw), and (told). The differences in the patterning of tokens were tested using two different non-parametric tests. The frequencies of vernacular tokens as they pattern within the three corpora are given in Table 13a and the frequencies of vernacular tokens as they pattern across the three corpora in Table 13b:

---

partially caused by the differences in collection methods and transcriptions.

Table 13a: Within group frequencies of vernacular forms\*

	<b>TLS</b>	<b>PVC</b>	<b>NECTE2</b>	<b>total</b>
<b>(can + NEG) N=260</b>	0% (N=0)	1.2% (N=1)	28.9% (N=33)	13.1% (N=34)
<b>(go) N=4567</b>	11.6% (N=84)	4.2% (N=93)	8.2% (N=132)	6.8% (N=309)
<b>(throw) N=86</b>	44.4% (N=8)	18.9% (N=7)	25.8% (N=8)	26.7% (N=23)
<b>(told) N=188</b>	6.7% (N=2)	14.3% (N=13)	7.5% (N=5)	10.6% (N=20)

\* the figures are proportions of vernacular forms out of the total number of tokens collected for that group. I.e. out of the total number of tokens collected for the speakers in the *NECTE2* corpus for the variable (can + NEG), 28.9% were vernacular forms. Similarly, out of the total number of tokens collected for the variable (can + NEG) across all three corpora, 13.1% were vernacular forms. The Ns are the total number of tokens collected for each variable.

Table 13b: Within variety frequencies of vernacular forms\*

	<b>TLS</b>	<b>PVC</b>	<b>NECTE2</b>	<b>total</b>
<b>(can + NEG) N=34</b>	0% (N=0)	2.9% (N=1)	97.1% (N=33)	100% (N=34)
<b>(go) N=309</b>	27.2% (N=84)	30.1% (N=93)	42.7% (N=132)	100% (N=309)
<b>(throw) N=23</b>	34.8% (N=8)	30.4% (N=7)	34.8% (N=8)	100% (N=23)
<b>(told) N=20</b>	10% (N=2)	65% (N=13)	25% (N=5)	100% (N=20)

\* the figures are proportions of vernacular forms in each corpus out of the total number of vernacular tokens collected for that variable. I.e. out of the total number of vernacular tokens collected for (go), 42.7% came from the *NECTE2* corpus. The Ns are the total number of vernacular tokens collected for each variable.

The table below summarises the descriptive statistics for the three subcorpora and the corpus as a whole for the Kruskal-Wallis test. The Ns given refer to the number of speakers included from each corpus.

Table 14: Summary of descriptive statistics for Kruskal-Wallis\*

<b>Corpus ↓ Variables →</b>		<b>(can + NEG)</b>	<b>(go)</b>	<b>(throw)</b>	<b>(told)</b>
<b>TLS</b>	N	20	37	17	18
	mean	0.00	0.11	0.46	0.05
	median	0.00	0.00	0.33	0.00
	SD	0.00	0.231	0.484	0.138
<b>PVC</b>	N	21	35	20	29
	mean	0.01	0.0498	0.21	0.11
	median	0.00	0.00	0.00	0.00
	SD	0.031	0.165678	0.386	0.310
<b>NECTE2</b>	N	24	42	14	29
	mean	0.29	0.10	0.08	0.06
	median	0.00	0.00	0.00	0.00
	SD	0.415	0.243	0.214	0.228
<b>FULL CORPUS</b>	N	65	114	51	76
	mean	0.11	0.09	0.26	0.08
	median	0.00	0.00	0.00	0.00
	SD	0.287	0.218	0.409	0.246

\* Kruskal-Wallis tests are based on ranked scores so it is not possible to talk about minimum and maximum scores as this value refers to the ranking score and not the proportional value on which the ranking is based.

I will present the results of both tests for each variable starting with (can + NEG). A chi-squared test of this variable found the difference between groups to be highly significant ( $X^2 = 45.032$ ,  $df = 2$ ,  $p$

< 0.001) and Cramer's  $V$  test for effect size (Pallant 2007: 217) revealed a highly significant medium effect (Cramer's  $V = 0.413$ ,  $p < 0.001$ ). A Kruskal-Wallis test further confirmed that there are significant differences between the distributions of vernacular tokens across the three corpora ( $p < 0.001$ ). As can be seen from the frequency distribution in Table 13b, *NECTE2* speakers contributed 97.1% of all vernacular tokens for this variable.

For the second dependent variable in this category, (go), chi-squared testing of the difference in frequency of vernacular forms between the three corpora showed a highly significant difference with a small effect size ( $X^2 = 56.618$ ,  $df = 2$ ,  $p < 0.001$ , Cramer's  $V = 0.111$ ,  $p < 0.001$ ). When looking at the frequencies given in Table 13b, we see that there is a steady increase in the number of vernacular forms used by the speakers with the highest proportion being used by the speakers from the *NECTE2* corpus. However, the Kruskal-Wallis test returned a clear, non-significant result ( $p = 0.288$ ).

The reverse seems to be the case for the third variable, (throw), where the chi-squared test found no significant differences between the three groups but the Kruskal-Wallis test did. The result of the chi-squared test was non-significant ( $X^2 = 4.049$ ,  $df = 2$ ,  $p = 0.132$ , Cramer's  $V = 0.217$ ), however, the result of the Kruskal-Wallis test shows a significant difference between the three groups ( $p = 0.034$ ). Comparing these results with the two frequency tables above (Tables 13a and 13b), it would seem that the use of vernacular *hoy* for Standard English *throw* is becoming more frequent again after a dip in the 1990s (PVC data).

The final variable is (told) and neither the chi-squared test ( $X^2 = 2.482$ ,  $df = 2$ ,  $p = 0.289$ , Cramer's  $V = 0.115$ ) or the Kruskal-Wallis ( $p = 0.691$ ) found any statistically significant differences between the use of vernacular forms across the three corpora.

To summarise, we can see that the two different tests sometimes yield different results, in this case for the two variables (go) and (throw). The implications of this are discussed further below. We can also see that the two tests yield similar results for two of the variables, (can + NEG) and (told), where (can + NEG) was found to be increasing in use over time and (told) to remain constant (i.e. there were no significant differences between the three groups). The table below summarises the results for all seven variables across the three statistical tests:

Table 15: Summary of inferential analyses for all seven variables

	(do + NEG)	(1 <sup>st</sup> pers.)	(2 <sup>nd</sup> pers.)	(can + NEG)	(go)	(throw)	(told)
Significant ANOVA	-	-	+				
Significant Chi sq				+	+	-	-
Significant K-W				+	-	+	-

To summarise, while not all seven variables showed significant differences in use over time, the ones that did (*youse* as second person pronoun, *canna* for sentential negation with *can*, *gan* for *go* and *hoy* for *throw*) all showed the highest frequency of use in the most recent data from 2007-2009. The results for *gan* and *hoy* are not as clear as those for *youse* and *canna*, as the two non-parametric tests reveal different results. One way to interpret this difference lies in the construction of the tests themselves. As discussed above, a chi-squared test reveals something about the difference in the patterning of individual tokens whereas the Kruskal-Wallis test is based on speaker ranks and thus compares the speakers across the three groups. In this way it could be argued that the data displays a significant difference in the amount of *tokens* of the vernacular forms of (go), however, the behavior of the speakers across the three corpora is not significantly different. This result could be caused by a general increase of use of all forms by all speakers in the newest data. In that case, the KW test would not find speakers to have changed their behavior significantly as the *proportion* of vernacular tokens for (go) would remain the same. Conversely, in the case of (throw) where the KW

test found a significant difference between the speakers' proportional use of standard and vernacular forms over time, the distribution of tokens across the three corpora was not significantly different according to the chi-squared test.

## 5. Discussion

A comparison of the two studies thus indicates that it is possible for the directionality of change in different levels of one variety to go in opposite directions. This highlights important issues in our understanding of phonology and morphosyntax but also raises questions as to which factors impact language change and how. I will here suggest that identity factors may explain how we can find levelling in the phonology of Tyneside English while at the same time see an increased use of local morphosyntactic forms.

### 5.1. *The difference between phonology and morphosyntax*

One suggestion which might help account for the difference in directionality of change is *salience* (Labov 1972, 1994; Trudgill 1986) as it can be hypothesised that if there are differences in the salience of certain phonological and morphosyntactic forms, or of these linguistic levels in general, it is probable that this will affect the direction or extent of language change. Specifically in the example of Tyneside English, it means that it is possible for there to be an increase in vernacular morphosyntactic forms but a decrease in vernacular phonological forms simultaneously, which we have seen above.

If the phonological level is more salient, speakers can deliberately aim for a more standard, or at least less local, accent or pronunciation whilst actually, subconsciously, increasing their use of vernacular morphosyntactic features. However, if the morphosyntax is more salient than the phonology then there can be a deliberate choice of vernacular morphosyntactic forms (e.g. to indicate local identity and group membership). The change in phonological forms will then be subconscious so that speakers might pronounce certain words in a more standard manner but actually opt for more overt vernacular features, such as the vernacular forms *yous* or *hoy*. At present this is purely hypothetical and it should be added that it is possible for different phonological forms to be more salient than others and similarly so for morphosyntactic forms and that this is also likely to vary between speakers.

A methodological issue arises, however, in the conceptualisation of 'more or less local' which is based on the very nature of phonology and morphosyntax, respectively. As we saw in Watt's (2002) study, a type of graded scale can be created for phonological variants to exemplify levels of 'localness' or 'standardness'. But it is not possible to create clines with intermediate values of different morphosyntactic variants of variables as these are binary in nature. For morphosyntactic variables only local forms (*yous*, *divn't*, *hoy*, etc.) and standard forms (*you*, *don't*, *throw*, etc.) exist.

It is important to keep in mind that, due to methodological differences, it is not possible to compare the two studies on a speaker-by-speaker basis. Even though data from the PVC corpus was used in both studies, proposing that speakers are using more vernacular morphosyntactic features at the same time as they are using less vernacular phonological features is not straightforward. None of the studies have looked at what each individual speaker is doing with regard to *both* phonology and morphosyntax. It is possible for Watt's results to show a change towards a regional standard (i.e. dialect levelling) and for the results of morphosyntactic study to find an increase in use of vernacular morphosyntactic forms. Based on the two studies reported here, we cannot be certain of how the changes are expressed by each individual speaker. In theory, at least, speakers can increase their use of vernacular forms on both the morphosyntactic level as well as on the phonological level; they can decrease their use on both levels; or simply increase on only one level (c.f. Cheshire et al. 2005).

## 5.2. *Linguistic local identity*

While differences in the salience of phonological and morphosyntactic variables may be one way to account for the different directions of change found in the *DECTE* corpus, speaker identity may also play a part. Watt (2002: 53) also comments on the role of a local linguistic identity and states that “[s]elf-identity in Newcastle, Tyneside and the north-east generally is rather complex” but adds that Tyneside inhabitants generally hold very strong feelings of regionality and local pride (2002: 54). He concludes by considering the role of the social changes of large industrial cities in the 20<sup>th</sup> century and its impact on the development of new regional identities (2002: 58) and states that this is a topic of interest across many disciplines, such as sociology, psychology, history and politics. The discussion which follows below touches on these aspects and suggests that socio-psychological factors (such as local affiliation) are linked to the salience of linguistic forms. In this way, social changes can impact speakers' language use through processes of social indexicality and enregisterment.

### 5.2.1. The urban regeneration of Tyneside

The Tyneside conurbation is an area which relied heavily on the ship-building and mining industries following the Industrial Revolution. In present times, however, these industries are no longer present on Tyneside and the area represents a “deindustrialised landscape” (Miles 2005: 913). The area was very much a working class area (and still is, to a degree) with the Quayside playing a central role both as an industrial centre but also as the connecting link between Newcastle and Gateshead. More recently, the Quayside has undergone a transformation. It is no longer the industrial centre it was but is instead a bubbling cultural area complete with an art museum, an award-winning bridge, several bars and pubs and a musical centre and so the overall function of the Quayside has changed from “production to consumption” (Miles 2005: 920). Miles (2005) links the development of and changes to the Quayside with the development of and changes in regional identity and quotes Wrightson (1995) who describes the Northern identity as being both closely tied to a sense of place but at the same time marred by an awareness of the questionable place of the location within the social landscape of England which seems to persist despite the recent 'culture-led regeneration' (Watt 2002). This problematic position of the North within England also spread from opinions about the inhabitants of the region to opinions about their language.

The stigmatisation of non-standard varieties is not unique to Tyneside. The ideology of the standard is “the belief that there is one and only one correct spoken form of the language, modelled on a single correct written form” (Milroy 1999: 174) and this leads to the stigmatisation of certain features and dialects as the most favourable opinions are most often linked to the standard language rather than to rural or urban dialects, for instance. In England, the RP accent is often perceived as constituting Standard (spoken) English and is often used as a point of reference for describing other varieties of English (Smith 1996: 66-67) and thus the social differentiation this implies has great impact on the perception amongst speakers of non-standard varieties, that is the broader local accents and dialects of the lower classes (Wells 1986: 14; Hughes et al. 2005: 15-16). It is clear that from the point of view of the standard ideology, then, any language change must be equated with a weakening of the language which carries with it further stigmatisation. This opinion can also be found among speakers of stigmatised varieties who view their own vernacular as inferior (Milroy 1999). Foulkes & Docherty (1999) and Kerswill (2003) suggest that in relation to the levelling found in the phonology of Tyneside English, young speakers find the traditional dialect old-fashioned and so opt for more modern pronunciations in order to sound more contemporary or less old-fashioned. This is not an abandonment of their northern background but rather just a move to sound modern. The new forms which are adopted must not be part of any other varieties as the young speakers still want to signal in-group membership with other speakers in their region.

We thus have two contradictions: one is the recent 'culture-led regeneration' of the Tyneside

conurbation focusing on more 'middle-class values' which seems to contradict the stereotype of 'backwards and working-class' which appears so persistent to both Tyneside speakers themselves but also to the rest of England. The second contradiction is the seeming divergence of change in the phonology and morphosyntax found in the *DECTE* corpus. When faced with the stigma attached to them and their local vernacular, Tyneside speakers have to decide whether they will let the stigma influence the way they speak, which is not necessarily a conscious or deliberate decision. Even though Tyneside English might be stigmatised in the rest of England (Watt 2002, Wales 2006), does not mean that it is so in the local speech community. It can be suggested that it is exactly the first contradiction which is part of the underlying cause of the second: with increasing globalization and loss of local identity in other areas, perhaps speakers are looking for ways to mark and express their identity.

For the sake of argument, it may be worthwhile to briefly consider an alternative to the contradiction hypothesis based on common perceptions in sociolinguistics as to the role of phonological and morphosyntactic variants. Phonology is probably the most well-researched area when it comes to language variation (Cheshire et al 2005 make a similar statement) and, as a legacy from Labov, no doubt, is often seen as the foremost indicator of personal identity. Conversely, variation in morphosyntax is often seen as a marker of style or register (although some argue that register shifts also encompass lexis and phonology, see Ferguson (1996 [1959]) on diglossia). Linguistic choices based on style are often the results of social background (class, education) but also to a large extent determined by speech situation, of course. Under this view, there is no contradiction in the directions of change found in the Tyneside data by Watt and me. On the contrary, the speakers' choice to sound northern, rather than Tyneside working class or southern, mirrors the urban cultural regeneration of Newcastle described above. The persistent or increased use by younger speakers of some local morphosyntactic forms is then a stylistic move which allows the speakers to keep a level of informality and friendliness (through use of non-standard forms) and Tyneside regionality. Thus, rather than diverging, the changes in phonology and morphosyntax could be seen as actually converging towards a place in between the local and the standardised, a place which reflects the values of a northern, post-industrial leisure-oriented society.

The issue with this alternative hypothesis, though, is that it assumes a number of things which we cannot necessarily take as fact. First of all, register or style can also be a badge of identity when a specific register is associated with a specific activity which denotes a community of practice (Lave & Wenger 1991; Eckert & McConnell-Ginet 1992). Thus, taking phonology as the primary index of personal identity cannot always be justified. Secondly, the area of variation and change in non-standard morphosyntax as well as comparisons between linguistic levels (such as phonology and morphosyntax) is an under-researched area (see Cheshire et al 2005). While it seems commonly accepted that speaker access to the different levels of language is not equal, how this difference manifests itself and the impact of it is not clear. I suggest that this taps into the notion of salience, a topic currently receiving a lot of attention from a variety of perspectives (Hollman & Siewierska 2006; Podesva 2011; Chiarcos et al 2011; Jaszczolt & Allan 2011; Rącz 2013; Jensen 2013). As this is thus a relatively young topic in linguistics, it is not possible to make strong claims about the connection between salience (which is related to both perception and production and has links to language learning, identity formation and styling, accommodation and other areas) and the role of speakers' access to or awareness of different levels in language.

### 5.2.2. The social value of local forms

As I mentioned in section 2.2, it has been proposed that linguistic identity is created through the linking of linguistic forms with cultural values through the process of enregisterment (Agha 2003). An understanding of this process and its link with the related concepts of social indexicality (Silverstein 2003) and indexical fields (Eckert 2008) lends further support to the argument that

social and linguistic issues are connected. Here, I adopt the view that these processes are furthermore linked with salience in that a linguistic form becomes salient exactly because of the attachment of social meaning. Salience is thus linked with indexicality as this is concerned with how social values emerge in a community and it is also linked with enregisterment as enregistered forms are forms where the social meaning of the forms *precedes* the original denotational meaning of the form. An elaboration of this argument can be found in Jensen (2013) but see also Honeybone & Watson (2013) for a similar argument based on data from popular dialect literature (which they call *Contemporary, Humorous, Localised Dialect Literature*) from Liverpool capturing the Scouse dialect.

Johnstone & Kiesling (2008) argue that different variables mean different things to different people and thus what constitutes sociolinguistic meaning then differs from person to person and is constantly in a state of flux. It is this web of socio-cultural meanings in a community which make up the indexical field of a linguistic form which (following Silverstein) speakers and hearers access when engaging in language. Thus we can see that considering factors such as enregisterment and social indexicality in the interpretation of patterns of language change can help shed light on the way in which Tyneside speakers anchor their identity when the place they have previously identified with is being changed and no longer represents values that are familiar to them (see above). This approach would account for the revitalisation (or steady use) of local vernacular forms to signify local identity.

Johnstone (2010) continues this line of argumentation and argues that the re-indexing of social meaning and enregisterment of vernacular forms (which she calls resemiotization) is caused by globalization (which I here interpret to be evident in Tyneside in the urban regeneration of the area) rather than being a reaction to it. The heart of the matter here is the very strong link between speakers' linguistic identities and their sense of place and belonging. In enregisterment processes of local vernacular linguistic features, it is precisely the social value and meaning of 'place' which is the primary index and, in turn, the local community which inhabit this place. This community, however, is a prime example of an *imagined community* (Anderson 2006) and the construction of a linguistic identity expressed through the use of a narrow repertoire of forms indexing this imagined community adds further to the *'them and us'* situation. This community is being constructed and reinforced by every use of an enregistered form, every display of local identity be it in written or spoken form, in casual conversation or for sale in the tourist office. It is this unbreakable link between language and place which is both built on cultural stereotypes and simultaneously feeds these stereotypes (which are in essence displays of *invented traditions*, (Hobsbawm 1992)) which makes this linguistic identity so powerful and important.

The results of the statistical analyses of the morphosyntactic variables revealed that the vernacular variants of all seven variables were either stable or increasing in use. This pattern could provide support for the idea of linguistic forms functioning as indexes of local affiliation and identity. However, we saw earlier that the same cannot be said for the broad Tyneside vowels in Watt's (2002) study, which he found were being replaced by less broad and more generally Northern forms. Even if Watt's conclusions point toward an abandonment of a very localised Tyneside identity in favour of a more general Northern identity, this difference in patterning is not necessarily a problem for the identity argument presented here due to a number of factors, the most prominent one being time. The reader is reminded that Watt's data was collected in 1994 whereas the data for the morphosyntactic study also included data from 2007-2009 and it is possible for there to have been a shift in the perception of the Tyneside conurbation within the area itself over the last two decades or so. Much of the recent urban regeneration mentioned in section 5.2.1 has taken place in the period after the PVC data was collected and, in recent years, media exposure of the Tyneside



area and language variety, such as the famous *Big Brother* narrator<sup>5</sup>, MTV's *Geordie Shore*<sup>6</sup> and TV presenters Ant & Dec<sup>7</sup>, has also become prominent on national TV in the UK. All of these factors may contribute to an increased sense of local affiliation and pride among the Tyneside inhabitants.

## 6. Conclusion

The results of the corpus study of changes in Tyneside English morphosyntax showed that vernacular variables were either stable or increasing. These results were compared to those of Watt (2002) who found that local diphthongs were levelling towards a regional standard. The difference in direction of change between these two was sought explained by a difference in the salience of the different levels of language linked to social changes through the processes of enregisterment and social indexicality.

A few caveats should be added, however. One issue concerns the quality and comparability of the data, mainly in the morphosyntactic study where data from three subcorpora is used, which was briefly mentioned in 2.2. This affects the reliability of the statistical results as they can only ever be as accurate as the data they are based on. Working with historical data means that researchers often have to make do with less than perfect data and very few ways to improve it. This is important to keep in mind in the interpretation of any results based on historical data even if all measures to even out the flaws have been taken. A second issue which was also mentioned above is the fact that the linguistic behaviour of the individual speakers was not mapped, i.e. even though the *PVC* data was used in both studies, the phonological and morphosyntactic tokens for each speaker were not compared on an individual basis. Finally, while several studies of language variation and speaker identity invoke arguments of salience, social indexicality, enregisterment and local affiliation as significant contributors to speakers' language use and changes over time (Llamas 2001; Burbano-Elizondo 2008; Beal 2009; Johnstone 2009; Honeybone & Watson 2013; Jensen 2013), further empirical work to test these supposed links is needed. Future studies of the connections between speakers' social experiences and psychological response mediated through language perception and use hold the potential for yielding valuable results which would help us understand the intricate web of connections which impacts on speakers' language use and thus language variation and change.

On a more specific level, the outcome of the comparison between the two studies based on *DECTE* data shows two things: the first is that speech corpora provide a valuable resource for studies investigating a variety of speech phenomena. We have seen that using the same data to investigate changes in different areas of a language variety has been fruitful and opened up new avenues for research. Secondly, this outcome further supports Cheshire et al. (2005: 167) who argue that just because levelling is found on the phonological level, we cannot expect to find it in other aspects of a language.

## References

- Agha, Asif (2003). 'The social life of cultural value'. *Language and Communication*, 23: 213-273.
- Anderson, Benedict (2006). *Imagined Communities*. New edn. London: Verso.
- Barnfield, Katie (2009). *Handbook to the Newcastle Electronic Corpus of Tyneside English [sic] 2 (NECTE2)*. <http://research.ncl.ac.uk/necte2/documents/handbook.pdf>. 27 January 2012.
- 5 The reality show *Big Brother* first aired in the UK in 2000 on Channel 4. The narrator is Marcus Bentley who was born in Gateshead on the southern bank of the river Tyne. <<http://www.marcusbentley.co.uk/>>
- 6 *Geordie Shore* first aired in 2011 and is a reality show set in Newcastle featuring North-East participants. It is currently on its eighth season. Marcus Bentley is also the narrator of this show. <<http://www.mtv.co.uk/geordie-shore/about>> <<http://www.marcusbentley.co.uk/>>
- 7 Anthony McPartlin and Declan Donnelly became famous after their roles in BBC's children's drama *Byker Grove* in the early 1990s and went on to record three multi-platinum albums. Since 1997 they have focused on presenting popular TV shows, such as *Pop Idol*, *Britain's Got Talent* and *I'm a Celebrity...* and awards ceremonies, such as *The BRIT Awards* and *Comic Relief*. <<http://www.officialantanddec.com/about>>

- BBC (2008). *Tyne*. URL: [http://www.bbc.co.uk/tyne/content/articles/2008/07/07/geordie\\_feature.shtml](http://www.bbc.co.uk/tyne/content/articles/2008/07/07/geordie_feature.shtml). Retrieved December 6, 2011.
- Beal, Joan (1993). 'The grammar of Tyneside and Northumbrian English'. In James Milroy & Leslie Milroy (eds.), *Real English: The grammar of English dialects in the British Isles*. Harlow: Longman. 187-213.
- Beal, Joan (2004). 'English dialects in the North of England: morphology and syntax'. In Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Meshtrie & Clive Upton (eds.), *A Handbook of Varieties of English – Vol. 2*. Berlin: Mouton de Gruyter. 114-141.
- Beal, Joan (2009). 'Enregisterment, commodification, and historical context: “Geordie” versus “Sheffieldish”'. *American Speech*, 84(2): 138-156.
- Beal, Joan (2010). *An Introduction to Regional Englishes*. Edinburgh: Edinburgh University Press.
- Beal, Joan, Lourdes Burbano-Elizondo & Carmen Llamas (2012). *Urban North-eastern English: Tyneside to Teesside*. Edinburgh: Edinburgh University Press.
- Bentley, Marcus (n.d.). *Marcus Bentley*. URL: <http://www.marcusbentley.co.uk/>. Retrieved August 14, 2014.
- Buchstaller, Isabelle & Karen P. Corrigan (2011). 'How to make intuitions succeed'. In April McMahon & Warren Maguire (eds.), *Analysing Variation in English*. Cambridge: Cambridge University Press. 30-48.
- Burbano-Elizondo, Lourdes (2008). *Language Variation and Identity in Sunderland*. PhD thesis: Sheffield University.
- Cheshire, Jenny, Viv Edwards & Pamela Whittle (1993). 'Non-standard English and dialect levelling'. In James Milroy & Leslie Milroy (eds.), *Real English: The Grammar of English Dialects in the British Isles*. Harlow: Longman. 53-94.
- Cheshire, Jenny, Paul Kerswill & Ann Williams (2005). 'Phonology, grammar, and discourse in dialect convergence'. In Peter Auer, Frans Hinskens & Paul Kerswill (eds.), *Dialect Change*. Cambridge: Cambridge University Press. 135-167.
- Chiaros, Christian, Berry Claus & Michael Grabski, M. (2011). *Saliency: Multidisciplinary Perspectives on its Function in Discourse*. Berlin: Walter de Gruyter.
- Corrigan, Karen P., Hermann L. Moisl & Joan Beal (2000-2005). 'A linguistic time-capsule: The Newcastle electronic corpus of Tyneside English'. URL <http://research.ncl.ac.uk/necte/>. Retrieved January 10, 2011.
- Corrigan, Karen P., Isabelle Buchstaller, Adam J. Mearns & Hermann L. Moisl. 2010-2012. *The diachronic electronic corpus of Tyneside English*. <http://research.ncl.ac.uk/decte/>. (5 March 2012).
- Croft, William A. (2006). 'The relevance of an evolutionary model to historical linguistics'. In Ole Nedergaard Thomsen (ed.), *Competing Models of Linguistic Change: Evolution and Beyond*. Amsterdam: John Benjamins. 91-132.
- Croft, William A. (2009). 'Toward a social cognitive linguistics'. In Vyvyan Evans & Stéphanie Pourcel (eds.), *New Directions in Cognitive Linguistics*. Amsterdam: John Benjamins. 395-420.
- Crystal, David (2009). *A Dictionary of Linguistics and Phonetics*. Hoboken, NJ: Wiley-Blackwell.
- Dobson, Scott (1978). *Scott Dobson's Geordie Recitations, Songs and Party Pieces*. Gosforth: Geordieland Press.
- Dobson, Scott (1986). *Larn Yersel' Geordie*. Morpeth: Butler Publishing.
- Dictionary of the Scots Language/Dictionar o the Scots Leid*. URL: <http://www.dsl.ac.uk>. Retrieved December 6, 2011.
- Eckert, Penelope (2008). 'Variation and the indexical field'. *Journal of Sociolinguistics*, 12(4): 453-476.
- Eckert, Penelope & Sally McConnell-Ginet (1992). 'Think practically and look locally: Language

- and gender as community-based practice'. *Annual Review of Anthropology* 21: 461-90.
- Ferguson, Charles A. (1996 [1959]). 'Diglossia.' *Word*, 51: 325-340. Reprinted in T. Huebner (ed.), *Sociolinguistic Perspectives: Papers on Language and Society*. Oxford: Oxford University Press. 25-39
- Foulkes, Paul & Gerard J. Docherty (1999). 'Urban voices – overview'. In Paul Foulkes & Gerard J. Docherty (eds.), *Urban Voices: Accent Studies in the British Isles*. London: Arnold Publishers. 1-24.
- Foulkes, Paul & Gerard J. Docherty (2006). 'The social life of phonetics and phonology'. *Journal of Phonetics*, 34: 409-438.
- Geeson, Cecil (1969). *A Northumberland and Durham Word Book*. Newcastle upon Tyne: Harold Hill.
- Geeraerts, Dirk (2005). 'Lectal variation and empirical data in Cognitive Linguistics'. In Francesco Ruiz de Mendoza & Sandra Peña Cervel (eds.), *Cognitive Linguistics: Internal Dynamics and Interdisciplinary Interactions*. Berlin: Mouton de Gruyter. 163-189.
- Geeraerts, Dirk (2010). *Ten Lectures on Cognitive Sociolinguistics*. Beijing: Foreign Language and Teaching Press.
- Gries, Stefan Th. (2009). *Quantitative Corpus Linguistics with R: A Practical Introduction*. London: Routledge.
- Hobsbawm, Eric (1992). 'Introduction: Inventing tradition'. In Eric Hobsbawm & Terence Ranger (eds.), *The Invention of Tradition*. Cambridge: Cambridge University Press. 1-14.
- Hollmann, Willem B. & Anna Siewierska (2006). 'Corpora and (the need for) other methods in a study of Lancashire dialect'. *Zeitschrift für Anglistik und Amerikanistik* 54: 203-216.
- Honeybone, Patrick & Kevin Watson (2013). 'Salience and the sociolinguistics of Scouse spelling: exploring the contemporary, humorous, localised dialect literature of Liverpool'. *English World-Wide*, 34(3): 305-340.
- Hughes, Arthur, Peter Trudgill & Dominic Watt (2005). *English Accents and Dialects: An Introduction to Social and Regional Varieties of English in the British Isles* (4<sup>th</sup> ed.). London: Hodder Educational.
- Jaszczolt, Kasia M., & Keith Allan (2011). *Salience and Defaults in Utterance Processing*. Berlin: Walter de Gruyter.
- Jensen, Marie M. (2013). *Salience in Language Change: A Socio-cognitive Study of Tyneside English*. PhD thesis: University of Northumbria.
- Johnstone, Barbara (2009). 'Pittsburghese shirts: Commodification and the enregisterment of an urban dialect'. *American Speech*, 84(2): 157-175.
- Johnstone, Barbara (2010). 'Indexing the local'. In Nikolas Coupland (ed.), *Handbook of Language and Globalization*. Chichester: Wiley-Blackwell. 386-404.
- Johnstone, Barbara & Scott F. Kiesling (2008). Indexicality and experience: Exploring the meanings of /aw/-monophthongization. *Journal of Sociolinguistics* 12(1): 5-33.
- Kerswill, Paul (2001). 'Mobility, meritocracy and dialect levelling: The fading (and phasing) out of Received Pronunciation'. In Pilvi Rajamäe & Krista Vogelberg (eds.), *British Studies in the New Millennium: The Challenge of the Grassroots*. Tartu: University of Tartu. 45-58.
- Kerswill, Paul (2003). 'Dialect levelling and geographical diffusion in British English'. In David Britain and Jenny Cheshire (eds.), *Social Dialectology*. Amsterdam: Benjamins. 223-243.
- Labov, William (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, William (1994). *Principles of Linguistic Change – vol. 1: Internal Factors*. Oxford: Blackwell.
- Lave, J. and E. Wenger (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge, Cambridge University Press.
- Llamas, Carmen (2001). *Language Variation and Innovation in Teesside English*. PhD thesis:

- University of Leeds.
- Miles, Steven (2005). 'Our Tyne': Iconic regeneration and the revitalisation of identity in NewcastleGateshead'. *Urban Studies*, 42(5/6): 913-926.
- Milroy, Leslie (1999). 'Standard English and language ideology in Britain and the United States'. In Tony Bex & Richard J. Watts (eds.), *Standard English: The Widening Debate*. London: Routledge. 173-206.
- MTV (n.d.). *Geordie Shore*. URL: <http://www.mtv.co.uk/geordie-shore/about>. Retrieved August 14, 2014.
- Official Ant & Dec* (n.d.). URL: <http://www.officialantanddec.com/about>. Retrieved August 14, 2014.
- Oxford English Dictionary Online, The.* (n.d.). 'Hoy, v.'. Oxford: Oxford University Press. URL: <http://www.oed.com/view/Entry/89050?rskey=AbfekI&result=7>. Retrieved July 31, 2013.
- Oxford English Dictionary Online, The.* (n.d.). 'Go, v.'. Oxford: Oxford University Press. URL: <http://www.oed.com/view/Entry/79544?rskey=raXKcZ&result=19&isAdvanced=true>. Retrieved December 6, 2011.
- Pallant, Julie (2007). *SPSS Survival Manual* (3<sup>rd</sup> ed). Maidenhead: Open University Press, McGraw-Hill Education.
- Podesva, Robert J. (2006). *Phonetic Detail in Sociolinguistic Variation: Its Linguistic Significance and Role in the Construction of Social Meaning*. PhD thesis: Stanford University.
- Podesva, Robert J. (2011). 'Salience and the social meaning of declarative contours: Three case studies of gay professionals'. *Journal of English Linguistics* 39(3): 233-264.
- R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Rácz, P. 2013. *Salience in Sociolinguistics. A Quantitative Approach*. Berlin, Boston: De Gruyter Mouton.
- Rowe, Charley (2007). 'He divn't gan tiv a college ti di that, man! A study of do (and to) in Tyneside English'. *Language Sciences*, 29: 360-371.
- Rowe, Charley (2009). 'Salience and resilience in a set of Tyneside English shibboleths'. In Stavroula Tsiplakou, Marilena Karyolemou & Pavlos Pavlou (eds.), *Language Variation: European Perspectives II*. Amsterdam: John Benjamins. 191-204.
- Silverstein, Michael (2003). 'Indexical order and the dialectics of sociolinguistic life'. *Language and Communication*, 23: 193-229.
- Smith, Jeremy (1996). *An Historical Study of English: Function, Form and Change*. London: Routledge.
- Trudgill, Peter (1986). *Dialects in Contact*. Oxford: Blackwell.
- Trudgill, Peter & Jack K. Chambers (1991). *Dialects of English: Studies in Grammatical Variation*. London: Longman.
- Upton, Clive, David Parry, & J.D.A. Widdowson (1994). *Survey of English Dialects*. London & New York: Routledge.
- Wales, Katie (2006). *Northern English*. Cambridge: Cambridge University Press.
- Watt, Dominic (2002). "'I don't speak with a Geordie accent, I speak, like, the Northern accent": Contact-induced levelling in the Tyneside vowel system'. *Journal of Sociolinguistics*, 6(1): 44-63.
- Watt, Dominic & Leslie Milroy (1999). 'Patterns of variation and change in three Newcastle vowels: is this dialect levelling?' In P. Foulkes and G. Docherty (eds.). *Urban Voices. Accent Studies in the British Isles*. Arnold Publishers: Great Britain. 25-46.
- Wells, John C. (1986). *Accents of English 2: The British Isles*. Cambridge: Cambridge University Press.
- Wright, Joseph (1898). *The English Dialect Dictionary*. London: Oxford University Press. URL:

<https://archive.org/stream/englishdialectdi03wriguoft#page/262/mode/2up>. Retrieved November, 2014.

Wrightson, Keith (1995). 'Northern identities: the longue durée'. *Northern Review*. 2: 25-34.