

5-1-2012

# "Ye ken fine wha I mean": Variation between you know and ye ken in Scottish varieties of English

Briony Jones

Follow this and additional works at: [https://digitalrepository.unm.edu/ling\\_etds](https://digitalrepository.unm.edu/ling_etds)

---

## Recommended Citation

Jones, Briony. ""Ye ken fine wha I mean": Variation between you know and ye ken in Scottish varieties of English." (2012).  
[https://digitalrepository.unm.edu/ling\\_etds/19](https://digitalrepository.unm.edu/ling_etds/19)

This Thesis is brought to you for free and open access by the Electronic Theses and Dissertations at UNM Digital Repository. It has been accepted for inclusion in Linguistics ETDs by an authorized administrator of UNM Digital Repository. For more information, please contact [disc@unm.edu](mailto:disc@unm.edu).

Briony Jones

*Candidate*

Department of Linguistics

*Department*

This thesis is approved, and it is acceptable in quality and form for publication:

*Approved by the Thesis Committee:*

Melissa Axelrod, Chairperson

Catherine Travis, Member

Graeme Trousdale, Member

**“YE KEN FINE WHA I MEAN”: VARIATION BETWEEN *YOU*  
*KNOW* AND *YE KEN* IN SCOTTISH VARIETIES OF ENGLISH**

**by**

**BRIONY ELIZABETH JONES**

**M.A. ENGLISH LANGUAGE,  
UNIVERSITY OF EDINBURGH, 2005**

**THESIS**

Submitted in Partial Fulfillment of the  
Requirements for the Degree of

**Master of Arts  
Linguistics**

The University of New Mexico  
Albuquerque, New Mexico

**May 2012**

## **DEDICATION**

This thesis is dedicated to Náná Tay. I think of you often.

## **ACKNOWLEDGEMENTS**

I would like to thank my parents for encouraging me to travel from an early age, and for supporting my decision to move to the United States to pursue graduate education. I extend my deepest thanks to Dr. Melissa Axelrod, chair of my committee, for her tireless encouragement of this project, and for guiding me closely from its inception to its completion. I would also like to thank the other members of my thesis committee: Dr. Catherine Travis, for her insights and thorough checking of my data; and Dr. Graeme Trousdale, for his encouraging comments on my work. Dr. Trousdale nurtured my passion for language when I was an undergraduate, and he left me eager to learn more. The support of my husband Glen has been invaluable in completing this project, and this task would have been much harder without his patience and understanding.

**May 2012**

**“YE KEN FINE WHA I MEAN”: VARIATION BETWEEN *YOU KNOW* AND *YE KEN* IN SCOTTISH VARIETIES OF ENGLISH**

**by**

**Briony Elizabeth Jones**

**M.A. English Language, University of Edinburgh, 2005**

**ABSTRACT**

The discourse marker *you know* is one of the most frequently studied discourse markers, and bears the hallmarks of a typical discourse marker with its complexity and indeterminacy of function (e.g. Holmes 1986, Erman 1987, Fox Tree & Schrock 2002, Irwin 2006). Much less work has been conducted into the uses of regional varieties of *you know*, specifically the use of Scots *ye ken*, partly as a consequence of a lack of data to study (Bauer 2004), and partly due to an assumption that there is little of interest to study in language varieties of Scotland (Murdoch, 1996).

In this thesis I present a discourse analysis of the variation between *you know* and *ye ken* as it is used in spontaneous conversations in Scottish varieties of English, taking my data from the SCOTS project, which aims to build a large electronic collection of the

languages of Scotland. I use Erman's (2001) monitor type framework to classify and categorize 270 tokens from 21 speakers across 7 recordings as functioning as one of three monitor types, and I look at the use of monitor types across two genres of conversation: one narrative, and one conversational. I also explore the roles that the sociolinguistic variables of age, gender, region of residence play in speakers' use of *you know* or *ye ken*.

I find that *ye ken* is preferred by older speakers and male speakers in more rural communities, while *you know* is preferred by younger speakers, women, and those in more urban communities. Amongst the eight speakers who use both *you know* and *ye ken*, use of *ye ken* is often primed by code-switching into Scots. *Ye ken* also functions as a marker of shared experience and reminiscence amongst older speakers in smaller, rural communities. The emergence of corpora of regional language varieties, such as SCOTS, enables this kind of study into the use of regional discourse markers, and provides the opportunity for comparative study of discourse markers across language varieties.

## TABLE OF CONTENTS

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Scotland in context: people, language, history</b>	<b>3</b>
2.1	Scottish language varieties	5
2.1.1	Scottish English	5
2.1.2	Scottish Standard English	6
2.1.3	Scots	6
2.1.4	Gaelic	7
2.2	The sociolinguistics of Scotland	8
2.3	<i>Ye ken</i> as a discourse marker in spontaneous conversation	10
2.4	Sources of data: the SCOTS project	12
2.5	Coding issues with SCOTS: orthography, transcription methods, and corpus limitations	14
2.5.1	SCOTS transcription methodology	14
2.5.2	SCOTS orthography: <i>ye</i> versus <i>you</i>	17
2.6	Speakers' demographic data	19
2.6.1	Speaker age	22
2.6.2	Speaker regions of residence	22
2.7	Summary	24
<b>3</b>	<b>Literature review</b>	<b>26</b>
3.1	Discourse markers	26
3.1.1.	<i>You know</i> as a discourse marker: meta-knowledge and manner of speaking	27
3.1.2	<i>You know</i> and domains of functioning	31
3.1.2.1	Textual monitors	32
3.1.2.2	Social monitors	34
3.1.2.3	Metalinguistic monitors	35
3.2	The role of audience: age, gender, and relationship between speakers	37
3.2.1	Age	38
3.2.2	Relationship between speakers	39
3.2.3	Gender	40
3.3	Genre and register	42
3.4	Summary	44
3.5	Research questions	45
<b>4</b>	<b>Methodology and Data</b>	<b>47</b>
4.1	Obtaining tokens	47
4.1.1	Narrowing the data and selecting tokens	47
4.1.2	Transcripts selected	48
4.1.3	Genre within the transcripts: narratives and conversations	51
4.2	Coding tokens	52
4.2.1	Coding for monitor type	52



4.2.1.1 Textual and metalinguistic monitors	54
4.2.1.2 Social monitors	56
<b>5 <i>You know, ye ken</i> and social groups</b>	<b>59</b>
5.1 Speakers' results by token type	59
5.1.1 Use of both <i>you know</i> and <i>ye ken</i> by speakers	60
5.2 Gender	61
5.3 Speaker age	64
5.4 Regional differences	66
5.5 Discussion	68
<b>6 <i>You know, ye ken</i> and domains of functioning</b>	<b>71</b>
6.1 Functional domains and monitor types	71
6.1.1 <i>You know</i> and <i>ye ken</i> as textual monitors	72
6.1.2 <i>You know</i> and <i>ye ken</i> as social monitors	78
6.1.3 <i>You know</i> and <i>ye ken</i> as metalinguistic monitors	81
6.2 Functional domains and genre	84
6.2.1 <i>You know, ye ken</i> and genre	88
6.3 Metalinguistic monitors: markers of shared experience	91
6.4 Discussion	95
<b>7 Conclusion</b>	<b>97</b>
<b>References</b>	<b>101</b>

## 1 Introduction

The focus of present study is an investigation into the distribution and function of the Scots discourse marker *ye ken* in talk-in-interaction, alongside the distribution and function of the standard English discourse marker *you know*. This study examines a number of factors which influence the selection of one form of this discourse marker over another amongst 21 speakers of Scottish varieties of English, many of whom use both *you know* and *ye ken* in spontaneous conversation.

There exists a substantial body of literature on the use of *you know* as a discourse marker in Englishes around the world (see Chapter 3), alongside a substantial body of literature on dialects and languages of Scotland. This project represents a convergence of the two research areas to produce a discourse analysis of the variation between *you know* and *ye ken* as it is used in spontaneous conversations in Scottish varieties of English in present-day Scotland. This kind of study is facilitated by the growing interest in corpora of non-standard varieties of language. Bauer (2004:107) notes that corpora have, for the most part, been built for national varieties of English and other languages rather than for regional dialects within one country. This is starting to change with the creation of corpora such as the Scottish Corpus of Texts and Speech (to be discussed in Chapter 4), which is the source of all data for this study, and which gives researchers a new and important opportunity to conduct systematic comparisons and analyses of established, and emerging, regional language varieties.

Following a literature review and the presentation of research questions in Chapter 3, Chapter 4 will focus on the methodological choices made for the present study, and presents speakers' demographic information. In Chapter 5 I discuss how *you*

*know* and *ye ken* are used by different social groups, and in Chapter 6, I discuss *you know* and *ye ken* and domains of functioning, building on a framework from Erman (2001). Chapter 7 presents a conclusion of the work.

## **2 Scotland in context: people, language, history**

According to the 2001 Scottish Census, Scotland's population is just over 5 million people, 25% of whom live in Scotland's largest cities of Glasgow, Edinburgh, and Aberdeen respectively. Only 4% of the population lives within the Highlands and Islands region of the country, which comprises over 790 islands in addition to parts of the mainland. Figure 1 presents a political map of the 32 council areas of present-day Scotland, and illustrates the high number of council areas in the Central region, in which the majority of the population resides. The Acts of Union 1707 joined the Kingdom of England and the Kingdom of Scotland together to form the United Kingdom of Great Britain, and Scotland has maintained a separate educational and legal system since that time, and also a separate Church of Scotland. Under the Scotland Act 1998, Scotland was granted limited self-government with its own parliament and elected representatives.

The present-day linguistic situation in Scotland is as rich, complex, and intriguing as the history of Scotland itself. Murdoch (1996:3) notes that it is “extremely hard” to establish the linguistic situation of Scotland: due in part to the national census only asking questions that relate to the speaking of Gaelic in the country, but also no doubt in part to what Auer refers to as an assumption that there is necessarily a natural relationship between geographical space and language variation (2005:3).

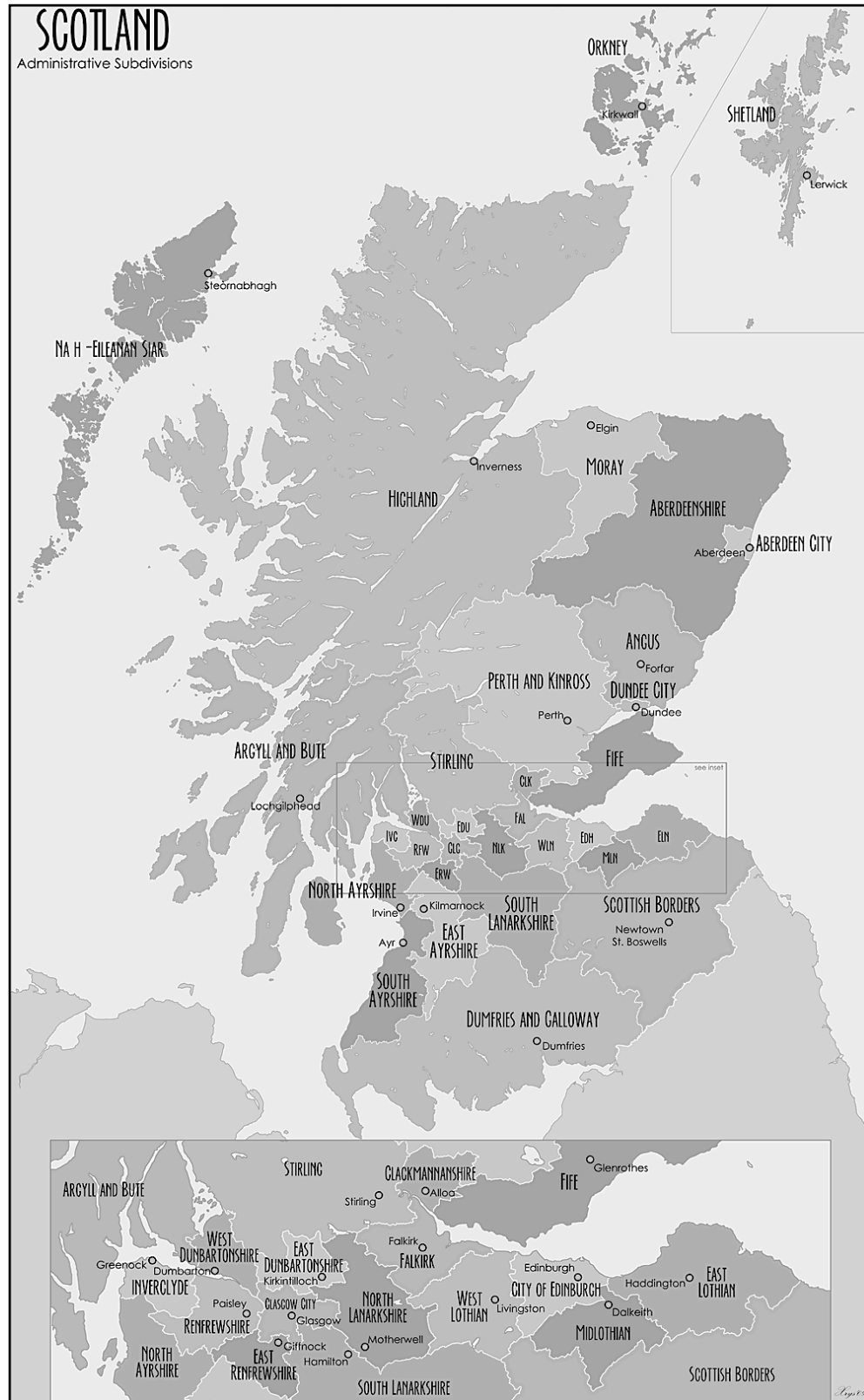


FIGURE 1. Map of the administrative divisions of Scotland, 2009.

## 2.1 *Scottish language varieties*

Though they may be hard, if not impossible, to exhaustively explain, the language varieties of Scotland are “a source of cultural wealth and pride to the nation” (Murdoch 1996:9). Speakers of Scottish English, Scots, Gaelic, and numerous other community languages make up the linguistic fabric of modern-day Scotland, with a dialect continuum existing not only between Scottish Standard English (SSE) and English Standard English, but also between regional and local varieties of SSE itself. One must also be aware of the differences between Scots and Scottish English, which, as Douglas (2003:23) notes, are both varieties “notoriously difficult to define.” Non-indigenous community languages such as Urdu, Punjabi, and Cantonese have also undoubtedly made their mark on the linguistic landscape of Scotland, and in ways that are only recently receiving attention from language scholars (Grant, 2000:41). Though the boundaries between language varieties are, and will remain, blurred, below I shall attempt to explain some of the more distinct features of each of these varieties.

### 2.1.1 *Scottish English*

Scottish English is a term for all native varieties in Scotland descended from Old English (Douglas, 2003:25), and thus has close historical links to English English<sup>1</sup>. Scottish English can be viewed as a localized form of British English “characterized by the inclusion of Scots lexis, Scottish grammatical and syntactic features, and Scottish pronunciation” (Douglas 2003:26). Scottish English also shares many features present in

---

<sup>1</sup> Per Trudgill (2002:2), English English refers to varieties of English as spoken in England, as distinct from varieties of English spoken across the United Kingdom as a whole.

English English lexis and grammatical and syntactic structures; however, Scottish English has a distinct phonemic inventory from English English, and phonological differences such as the Scottish Vowel Length Rule (SVLR), in which a vowel is long before a voiced fricative, /r/, or a morpheme boundary (#), but is otherwise short. Within the category of Scottish English, we have Scots and Scottish Standard English, but also a number of varieties that are intermediate between the two.

### 2.1.2 *Scottish Standard English*

Scottish Standard English (SSE) describes the standardized variety of Scottish English, which is in widespread use across Scotland. It is understood by the vast majority of the population and is used in media broadcasts, schools, business, and government, giving it a prestige status. McClure (1994:79) identifies SSE as being characterized “most obviously by pronunciation”, and that regional, social, and age and gender variations do exist, principally on the phonetic level and “very noticeably in intonation”.

### 2.1.3 *Scots*

The term *Scots* is used to describe a wide range of language varieties past and present in Scotland. The definition encompasses numerous regional and social dialects (Douglas 2003:26), and Douglas notes that it is “generally held” to include localized vernaculars known as broad Scots or dialect Scots, such as Ayrshire and Doric dialects, but also contemporary urban dialects such as Glaswegian, usually spoken by urban working class populations. Another variety of Scots, Lallans, was originally devised as a “standard” literary language, and incorporates vocabulary from many Scots varieties, as

well as incorporating words from old Scots dialects that are now obsolete (Douglas 2003:26). *Ye ken* could be considered a Scots phrase, given its association with literary varieties of Scots (see Welsh (1998) in 2.3 below), and with what Aitken (1979:107) identifies as an “overt Scotticism [...] that special diction of Scottish-tagged locutions used self-consciously by many Scottish speakers as a kind of stylistic grace and as a way of claiming membership of the in-group of Scotsmen.” It is easier to identify *ye ken* as being a part of Scots than of SSE, but as Scots is a very broadly defined variety, this label alone does not tell us much about the use of *ye ken*.

Despite its strong representation in Scottish literature and dramatic arts and its frequent invocation for anything wishing to be defined or viewed as regional, McEwan-Fujita (2011:57) notes that neither Scots nor Gaelic has played any significant role in nationalist politics in Scotland up to the present day. Trudgill (2000:144) makes note of the attempt of some communities to re-establish Scots as a language in its own right. Murdoch (1996:28) notes that cultural issues become political ones when they form an important expression of national identity, and provides evidence to illustrate that the issue of Scots language endorsement and protection may become a political issue in the future.

#### 2.1.4 Gaelic

Though not the focus of this study, Gaelic or Gàidhlig is a variety from which many modern Scots words originate. 58,000 Scottish residents are reported to be speakers of Gaelic (2001 Census), in contrast to the 65,978 speakers identified in the 1991 Census (MacKinnon, 1991:106). 20% of Gaelic speakers in Scotland live in the Gàidhealtachd, the traditionally Gaelic speaking areas of the Highlands and Islands (2001 Census). In a



discussion of the western retreat of the Gaidhealtachd, MacKinnon notes that numbers of Gaelic speakers have rapidly declined among the young (1991:108). The Gaelic Language (Scotland) Act 2005 gives formal recognition to Gaelic for the first time, establishing the body Bòrd na Gàidhlig to promote the use and understanding of Gaelic language and culture (Gaelic Language (Scotland) Act 2005 (asp 7):1).

As with most of the world's indigenous languages, Gaelic is endangered, and efforts of language maintenance can be witnessed through bilingual signs at train stations, the publishing of government information in both languages, and BBC television programming in Gaelic for children. As is often the case when bilingualism is promoted as policy, some Scottish residents are skeptical of the push to re-introduce Gaelic into the everyday environment: McEwan-Fujita notes “In late 20th-century Scotland, the increased visibility of Gaelic generated complaints from some journalists about the unpleasant burden of exposure to Gaelic as a foreign language. This seemingly arises from an ideological-affective complex of anti-Gaelic prejudice and a belief in the onerousness of bi- or multilingualism” (2011:58). Murdoch states that it is not uncommon for scholars in certain fields to assume that, if only 2% of Scotland speaks Gaelic, 98% must speak a homogeneous variety of English, possibly even Standard English as it is spoken in England (with no acknowledgement of England’s myriad dialects).

## 2.2 *The Sociolinguistics of Scotland*

In the 16th Century, Scots as a variety was still fairly autonomous from Midland and South-Eastern English. Aitken (1979:90) proposes that Scotland during this era did

not have a great patriotic objection to the infiltration of English spelling, lexis, and grammar into the Scots language, and that contemporary documents support this idea that Scots was viewed as a “homely, domestic, maternal language” – not one suited to the gentry. Pronunciation, however, was a different issue, with Scots pronunciations being retained into the 17<sup>th</sup> Century over their English counterparts. By the time of the Union of Parliaments, the language of the gentry of Scotland approximated to English Standard English (1979:95), and in a reflection of what Aitken terms “linguistic insecurity and self-consciousness”, beginning in 1752, a series of lists of Scottish words and expressions was published, for Scots to learn and purge from their own language variety (1979:96).

The 19<sup>th</sup> Century saw a change in attitude, however, with the increasing popularity of Scots Romantic writers such as James Macpherson, Walter Scott, and Robert Burns. During this period, there was a reduced desire to extinguish Scots as a language, and instead a drive to connect Scots with a growing nationalistic sentiment. As Auer (2005:3) notes, the origins of systematic study of dialect emerged during the late 19<sup>th</sup> Century, at a time when nationalist thinking in Europe was at its climax.

Aitken (1979:98) quotes from *The Second Statistical Account of Scotland* of 1845, whose anonymous author stated “The most common dialect is a mixture of Scotch and English, the Scotch used being of the somewhat vicious kind, known, I believe, by the name of Aberdeenshire. The Scotch, however, is dying out. Every person remembers the frequent use, in former years, of terms and phrases that are now seldom heard but among the older and more secluded”. At this point in history, many value judgments were passed on the type of Scots spoken in various locales, with R. de B. Trotter, writing in 1901, discussing the “wonderful gibberish which now passes current for Scotch” (quoted in

Aitken, 1979:98). Aitken also cites a Scottish Education Department report from 1952, which states that the Scots to be spoken in the classroom should be that found to be generally acceptable by educated Scots, but that “slovenly perversions of dialect will be excluded” (Aitken, 1979:98).

Trudgill (2000:50) touches on the issue of situational switching as it happens between Scots dialects and SSE. Speakers of Scots dialects may switch, in more formal situations, to SSE, and Trudgill provides the examples of a switch from *I’ve kenned yon man eight years* to the SSE *I’ve known that man eight years*, and notes that the jump between the two varieties is quite considerable. For children, it requires the learning of SSE lexical items and pronunciations, most of which happens in a school context if SSE is not spoken in the home. This situational switching between varieties is at the center of the present study, and speakers featured in the recordings used for this study frequently switch between different varieties of Scottish English during talk-in-interaction, suggesting that such a phenomenon may be widespread.

### 2.3 Ye ken as a discourse marker in spontaneous conversation

*Ken* ‘know’ is an example of a verb that would be used primarily in a home context. Smith, Durham & Fortune (2007) conducted a cross-sectional analysis of the acquisition of variable forms in a Scottish dialect by young children, and provided the following examples of *ken* in use in conversation between young children and their adult caregivers:

- (1) Caregiver: You do na **ken** how to stand still.
- (2) Lucy, 3;4: How? How can I nae go in the washin’ machine?

Caregiver: Cos it'll maybe go- get switched on. And then you'll be stuck in it.  
And you'll droon!

Lucy, 3;4: Fit?

Caregiver: You'll droon.

Lucy, 3;4: I'll go awa', won't I? You'll be lookin' far your- your Lucy is at, won't you? And you winna **ken** far to get to my clais, will you?

Caregiver: No, I'll be greetin'.

(From Smith, Durham & Fortune, 2007)

In examples (1) and (2), *ken* is being used, by both children and caregivers, in its canonical sense of “to make known, declare, confess, acknowledge” (Oxford English Dictionary). This usage was first attested in 1219 in *Beowulf* and is derived from Old Saxon *kennian* and Middle Dutch *kennen*. A later usage of *ken* attested since the 1300s is “To know, understand, or perceive (a fact, etc.); to be aware of, to be aware *that* (*what*, etc.).” This form appears in Dickens's *Christmas Carol* as well as other contemporary literature. The Oxford English Dictionary notes that *know* is generally held to be from the same root as *ken*, and it worth noting *ken*, *can*, and *know* are all etymologically related. Both *know* and *ken* have become polysemous over time, with both now used as a discourse marker in the form of *you know* and *ye ken* respectively. The discourse marker *you know* has various uses, but is most commonly used to structure a conversation, monitor turns, and establish shared meanings with an audience (see Chapter 2).

The lack of existing literature on the functions and use of *ye ken* as a discourse marker means that its most common uses in discourse are unknown; however, in (3) below, we see an example of *ken* functioning as a discourse marker utterance finally. In a sketch from Scottish BBC television comedy called *Chewing the Fat*, titled “Taysiders in Space”, we are presented with a Starship Enterprise manned entirely by residents of the former local government region of Tayside in western Scotland:

- (3) Comms: Captain!  
 Captain: What noo?  
 Comms: There's a great big hoo-er o a spaceship comin' towards us; ye want tae see the bastard size o the thing, it's a guid yin or twa sizes bigger than oors, **ken**!  
 (Chewing the Fat: "Taysiders in Space")

*Ye ken* can be used with a null subject to be simply *ken*, as in (3) above<sup>2</sup>. However, for the present data set, the form *ye ken* was used for study, as in (4) below, in which a woman explains that she did not hear her parents refer to each other by their first names, and (5), an excerpt from the novel *Filth* by Scottish novelist Irvine Welsh:

- (4) F1043: It was a lang time right enough before I ever heard the first names  
 //mentioned, **ye ken**?//  
 (1448: BBC Voices Aberdeen)
- (5) The thing is Ray, you've earned a lot ay respect in this department, and it's starting tae get tae the auld boy. **Ye ken** what ah'm saying? I look Lennox in the eye. He's getting the drift I want him to get. It's the young stag syndrome.  
 (Welsh, 1998:195)

The focus of the present study is a discourse analysis of the use of this regional discourse marker, *ye ken*, in talk-in-interaction, alongside a discussion of sociolinguistic factors that may influence the use of this form. In Chapter 3, I present an overview of the literature on the use of *you know* in talk-in-interaction, from which there is a substantial body to draw.

#### 2.4 Sources of data: the SCOTS project

Data for the present study were taken exclusively from the University of Glasgow's Scottish Corpus Of Texts & Speech (SCOTS), which has been available on the

---

<sup>2</sup> Aitken (1979:109) claims that *ye ken* carries less "stigma" than *ken*, but does not provide anything to back up this assertion. As *ken* with the null subject does not appear within the transcriptions used for the present work, it is hard to make further comment.

internet in an open access form since 2004. The corpus can be accessed by anyone with a computer and at any time, making it easy to use in classrooms, libraries, public institutions, and private homes. SCOTS is the work of Glasgow's English Language Department and the STELLA project (Software for Teaching English Language & Literature and its Assessment) of the School of English and Scottish Language and Literature at the University of Glasgow. As Fiona Douglas notes in an article on the design of SCOTS, the aim of the project is to build a large electronic collection of both written and spoken texts of the languages of Scotland (2003:23). As of December 2011, the corpus contains 4,286,620 words. There are over 1,190 texts in the corpus, of which 80% are written and 20% spoken. Most of the spoken texts were recorded since the year 2000, and specifically for the SCOTS project (SCOTS, "Details"). SCOTS aims to cover the period from 1945 to the present day, though it is noted on the SCOTS website that the majority of texts are from the latter part of this era.

As Douglas notes, the SCOTS project is the first large-scale project of its kind for Scotland (2003:23). Although other Scottish corpora do exist, there has never been an electronic archive specifically dedicated to the languages of Scotland, and the SCOTS corpus aims in its later phases to include not only Scots and Scottish English but also Gaelic and non-indigenous community languages such as Punjabi, Urdu, Italian, and Chinese (Douglas 2003:24). Douglas also notes that interest in constructing corpora for languages other than Standard English is increasing, and that "many scholars are now turning their attention to the peculiar problems presented by attempting to construct well-balanced representative corpora for such language varieties" (2003:24). Douglas notes that SCOTS is intended as a publicly available resource, and that for several reasons such

a corpus is necessary and perhaps also long overdue. Douglas argues that “surprisingly little” reliable information is available on a variety of linguistic issues such as the survival of Scots; the distinguishing characteristics of Scottish English; or the use of the non-indigenous community languages (2003:24). In addition to providing a resource for answering such questions, SCOTS will also function as a way of persevering threatened languages such as Gaelic for future generations (2003:24).

## *2.5 Coding issues with SCOTS: orthography, transcription methods, and corpus limitations*

Once the seven texts to be used in this study were identified, I began to select and code the individual tokens of *you know* and *ye ken*. In addition to coding for genre, I also coded tokens for document number; speaker ID number (a four digit number preceded by F for female speaker and M for male speaker); speaker gender; speaker decade of birth; speaker geographic location; the utterance context in which the token occurred; and monitor type (see discussion in Chapters 3 and 4). Some categories were more complicated to code for than others, for the reasons discussed below.

### *2.5.1 SCOTS transcription methodology*

The “Details” page on the SCOTS website explains some of the transcription methodology used for the recordings, specifically the following (with language not changed from the website description):

- Overlaps: //
- Censored material: [CENSORED: description of material censored]
- Inaudible: [inaudible]

- Unclear: [?]part of transcription that transcriber is unsure about[?]
- False starts: false starts, stammering and truncated words are tagged and appear in the transcription followed by a hyphen [nineteen f-f-f-fifty nine]
- Semi-lexical items: appear unmarked in transcription, but are tagged in the underlying form [er not until I was kind of older uh huh]
- Non-lexical items (paralinguistic information): [cough], [laugh]
- Non-linguistic events: [phone rings], [plate is dropped]

The SCOTS “Details” page also mentions that transcriptions of audio recordings were made using Praat in order to be able to time-stamp and synchronize transcriptions with their recordings: one useful feature of the SCOTS texts is the ability to hear the conversation recording play in sync with the transcription of the conversation on the website, with a speaker’s turn highlighted on screen as the recording moves through the conversation.

However, it does not appear that Praat was used, at least in any consistent way, to demarcate intonation units in the transcripts. In discussing the intonation unit, Chafe (1994:69) says “spoken language lends itself to segmentation into intonation units. Such units are identifiable on the basis of a variety of criteria, among which are pauses or breaks in timing, acceleration and deceleration, changes in overall pitch level, terminal pitch contours, and changes in voice quality.” He notes that these intonation units can “[open] up investigations of both language and the mind by observing language as it really is”, and that observations of naturally produced speech and writing can help us understand certain important ways in which language and consciousness interact (1994:301—302). These properties make the intonation unit a valuable one for the study of talk-in-interaction. However, such an option is not currently available for transcriptions from audio recordings featured in SCOTS. Where commas, exclamation



marks, periods, and question marks appear in SCOTS transcripts, they seem to be fulfilling more of a textual purpose for the reader than they do of marking any intonational characteristics of the speakers' utterances. Occasionally a comma seems to indicate a continuing intonation contour, but it can just as likely indicate that a transcriber felt a comma was necessary for ease of reading the transcribed recordings. As an example, I have presented M1013's utterance in (6) as if the punctuation used within the transcript were to indicate four separate IUs, retaining the original punctuation of three commas and one period.

- (6) M1013: Bert [CENSORED: surname] an I've just reached my eightieth birthday,  
an I've been in Hawick aw my life,  
except when I was in the forces just,  
that's it.

(1430: BBC Voices Hawick)

In listening to the recording, it is apparent that the first IU ending with a comma does indeed have a continuing contour, but the second IU, transcribed with a comma, has a final contour. The third IU, "except when I was in the forces just," is not really a separate IU at all, as there is nothing but a 0.3 second pause between this and the start of the fourth IU, "that's it", which does itself have a final contour. It appears that a comma has been placed after "in the forces just" in order to increase readability of the transcript. A period featured in a transcript does sometimes indicate a final intonation contour, as in (28), but in some cases it is, again, used for ease of reading and to break up a narrative portion in the transcript. Based on my listening to the recordings, on occasions where a question mark is featured, it does seem to indicate an appeal contour. However, given the unreliability of punctuation to mean more than just a structuring of the transcribed

discourse for the reader's benefit, I will refer to punctuation as a discussion point and nothing more.

The intonation unit was not selected as a basic unit for this study, for the simple reason that intonation units were not transcribed in the data that I used, and coding for this would be beyond the scope of this study. Additionally, the focus of this study is the categorization of tokens by monitor type, and Erman's categorization of monitor type does not focus on the intonation unit. The lack of coding for intonation units in transcripts from this corpus does raise questions as to what kinds of transcriptions should be available for a database such as SCOTS, which aims to cater to a scholarly audience alongside many other audiences. As the compilation of corpora of non-standard language varieties is still in its emergent stage, and as the data available within SCOTS is continuing to grow, hopefully this is an issue that can be addressed in the future.

### 2.5.2 *SCOTS orthography: ye versus you*

The team of transcribers for the SCOTS recordings did not use any standard spelling method in their transcripts. As Douglas explains: "Spelling in Scots is not standardized, although the Scottish Language Dictionaries organization is currently working on establishing preferred spellings. The same word may be spelled a variety of ways [...] Thus, there are problems when devising a lexical search system if trying to search for what is essentially the same word across different texts" (2003: 31). The SCOTS website explains:

"The SCOTS Corpus contains documents in Scottish Standard English, documents in different varieties of Scots, and documents which may be described

as lying somewhere between Scots and Scottish Standard English. While Scottish Standard English has a standard written form, Scots does not. This means that the corpus contains a wide range of variation in spelling. We hope to offer a means of searching for all of the variant spellings automatically in the future. In the meantime, we recommend the online *Dictionary of the Scots Language* as an excellent source of possible variants.” (SCOTS, “Details”)

The website also notes that “[t]he general notes above on spelling and variation also apply to transcriptions of spoken material. In addition, with transcribed spoken documents, it was necessary to decide upon conventions in order to make the transcriptions as consistent as possible [...] Where the dictionary offers alternative spellings for a word, the one closest to the speaker’s pronunciation is selected” (SCOTS, “Details”).

The issue of not using a standardized spelling method became apparent to me when searching for transcripts featuring *you know* and *ye ken*, and coming across these tokens alongside tokens transcribed as <ye know> and <you ken>. These tokens appeared not only within the same transcript, but sometimes from the same speaker. As an example, M1015 in (7) produces three tokens, transcribed as *ye ken*, *you know*, and *ye know*:

- (7) M1015: But eh jocularly I say "Is the b-", "Have you brought your brat wi ye?", **ye ken**, but that's it as a just ehm a jokin thing but that is if it just as Kenny says if they've been bad, **you know**, it's eh it's just bein a little brat, **ye know**, so it it carries on.

(1431: BBC Voices Dunbar)

It was for this cited reason of “a wide range of variation in spelling” used in the SCOTS transcripts that <ye know> was not considered a separate token type from <you

know>; nor was <ye ken> considered a separate token type from <you ken>, and therefore the tokens for this study include all four of these variants. When listening to each of the tokens, I was able to hear that there was indeed quite a variation in transcriber interpretation of a speaker's pronunciation, despite clear attempts to be as consistent as possible. There were pronunciations of *ye* and *you* from across the phonological spectrum, from more closed front vowels and back vowels to more open back vowels, and also mid-central vowel pronunciations. This is to be expected with a lexical item containing a vowel that frequently undergoes coarticulation and reduction in spoken discourse.

In some cases I concurred with the transcribers' representation of the second person pronoun as <ye> instead of <you> for example, but in other cases I did not, or I thought it should have been represented as something else such as <y-> (I did not find any tokens transcribed with <y->, <y'>, <ya>, or anything similar). Thus, I retained the original transcriptions of either <you know>, <ye know>, < you ken>, or <ye ken>, and maintained my focus on *ken* versus *know*: two lexemes which are much harder to confuse. Only one instance of <you ken> appeared in this study; all other *ye ken* tokens were transcribed as <ye ken>: an example of transcribers attempting to be as consistent as possible, although once again, the pronunciations within the recordings did not always match with what was transcribed. As <ye ken> is by far the most commonly transcribed form in the records, I choose to refer to *ye ken* instead of *you ken* throughout this study. The lack of a standard spelling for many other dialect items used in the recordings did not pose any notable problems for this project.

## 2.6 *Speakers' demographic data*

The speakers whose conversations provided data for the SCOTS project, and subsequently for this project, provide some interesting information of their own in the form of demographic information. For all speakers but one who feature in this study, SCOTS provided speaker decade of birth; place of residence; gender; occupation; educational level reached; religious affiliation; parents' regions of residence and occupations, amongst many other categories. As mentioned in 4.3.2., I gathered demographic data for each speaker detailing speaker decade of birth, geographic location, and gender, in order to keep the demographic variables to a reasonable number, and also to reflect three of the most commonly investigated variables in sociolinguistic research. Table 1 summarizes each speaker's decade of birth, gender, and the genre of recordings in which they appear.

Data was not available for speaker F1054's decade of birth, nor her location: F1054 appears in all four of the BBC Voices transcripts, and provides tokens of *you know* or *ye ken* in three of these transcripts. The role of F1054 in the featured BBC Voices recordings is that of interviewer, tasked with eliciting dialect forms from local residents who are participating in the recordings. Based on the information provided on the BBC Voices website, F1054 is most likely a radio broadcaster for one of BBC Radio's four Scottish stations, and this status makes it very unlikely that she is a juvenile. Beyond this, however, little can be stated for certain about F1054's demographic information, other than her gender: from listening to the BBC Voices recordings, it is apparent that she is female.

TABLE 1. Speakers' demographic information.

Speaker #	Decade born	Gender	Transcripts featured in
<b>606</b>	1940s	F	819, Conversation about family life
<b>646</b>	1920s	F	351, Conversation about childhood memories
<b>822</b>	1940s	M	819, Conversation about family life
<b>823</b>	1940s	F	819, Conversation about family life
<b>826</b>	1940s	F	1042, Conversation about dialects
<b>902</b>	1940s	F	1042, Conversation about dialects
<b>903</b>	1940s	M	1042, Conversation about dialects
<b>1004</b>	1940s	M	1428, BBC Voices: Glasgow
<b>1006</b>	1940s	F	1428, BBC Voices: Glasgow
<b>1010</b>	1930s	M	1430, BBC Voices: Hawick
<b>1011</b>	1920s	F	1430, BBC Voices: Hawick
<b>1012</b>	1950s	M	1430, BBC Voices: Hawick
<b>1013</b>	1920s	M	1430, BBC Voices: Hawick
<b>1014</b>	1930s	M	1431, BBC Voices: Dunbar
<b>1015</b>	1940s	M	1431, BBC Voices: Dunbar
<b>1016</b>	1970s	M	1431, BBC Voices: Dunbar
<b>1040</b>	1920s	F	1448, BBC Voices: Aberdeen
<b>1041</b>	1930s	F	1448, BBC Voices: Aberdeen
<b>1042</b>	1930s	M	1448, BBC Voices: Aberdeen
<b>1043</b>	1930s	F	1448, BBC Voices: Aberdeen
<b>1054</b>	--	F	1428, BBC Voices: Glasgow; 1430, BBC Voices: Hawick; 1431, BBC Voices: Dunbar
<b>TOTAL</b>	--	<b>11F/10M</b>	

F1054 does produce 17 tokens used in the present study, 15 of *you know* and 2 of *ye ken*, and so her data is important to include: she produces the fifth highest number of tokens from a single speaker. Moreover, her role as mediator, elicitor, and chair in the BBC Voices recordings is rather central as we shall see, and thus her data was important to

include. Her lack of available demographic data is factored into all subsequent calculations in which demographic variables other than gender are measured.

### *2.6.1 Speaker age*

From a total of 21 total speakers featured in this study, 9 were born before the 1940s and 11 were born during or after the 1940s, with speaker F1054's decade of birth being unknown. The recordings used in this project were all made between 2002 – 2005, and as such, we can assume that speakers born in the 1940s were aged between 53 – 65 at the time of recording. Many of these speakers would be considered “baby boomers”, born into an era of massive post-war changes in British economy and society, including the emergence of a more transient population who frequently relocated to other towns and cities for work. Giddens (1997:49) notes that “the populations of inner cities have declined as inhabitants move out to the suburbs and dormitory towns”, and that these shifts in demographic patterns over the last forty years have been viewed by some as the possible source of an explanation for the dialect leveling that is occurring throughout much of the British Isles.

Nine of the 11 speakers born in or after the 1940s were born in the 1940s itself. The youngest speaker, born in the 1970s, was aged between 26 – 35 at the time of recording, and the oldest speakers, four of whom were born in the 1920s, were aged between 73 – 85 at time of recording. This project provides a snapshot of roughly 59 years of synchronic variation in Scottish English from which data can be analyzed.

### *2.6.2 Speaker regions of residence*

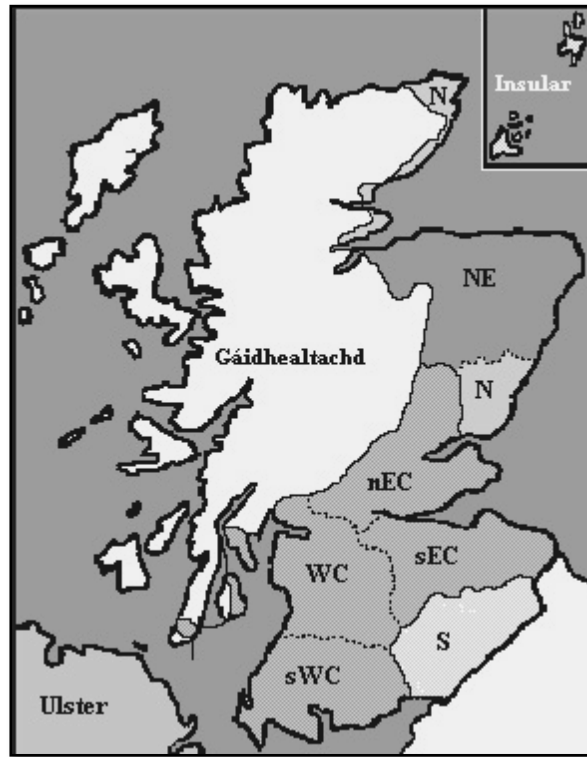


FIGURE 2. Dialect Map of Scotland (Eagle, 2001).

To simplify the results from the data, the dialect areas for the present study were collapsed into the following regions: Southern Scots (S in Figure 2); Northeastern Scots (NE in Figure 2); Northern Scots (N in Figure 2); and Central Scots (sWC, WC, sEC, and nEC in Figure 2). Tokens were obtained from speakers in every Scottish dialect region except the Gáidhealtachd, and so, in an unsurprising result, zero tokens from the Gáidhealtachd region are featured. Of the total of 20 speakers for whom region of residence information is available, seven speakers were born in and live in the Central Scots dialect region, which goes some way to explaining why just over half of all tokens come from speakers in the Central Scots dialect region. Table 2 provides a further breakdown of place of residence for these 20 speakers, and provides the name of the towns or cities in which they live. The cities of Glasgow and Aberdeen have populations of around 600,000 and 220,000 respectively. The town of Hawick has a population of



under 15,000 residents, while the towns of Torry and Dunbar have fewer than 10,000 residents. The towns of Whitemyers, Auchenblae, and Chryston have fewer than 5,000 residents.

TABLE 2. Speakers' town and region of residence.<sup>3</sup>

	<b>Number of speakers</b>	<b>Totals by region</b>
<b>Northeastern Scots</b>		
Whitemyers	2	
Torry	2	
Aberdeen	2	<b>6</b>
<b>Northern Scots</b>		
Auchenblae	3	<b>3</b>
<b>Central Scots</b>		
Dunbar	3	
Chryston	1	
Glasgow	3	<b>7</b>
<b>Southern Scots</b>		
Hawick	4	<b>4</b>
<b>TOTAL</b>	<b>20</b>	<b>20</b>

## 2.7 *Summary*

The SCOTS project proved to be an invaluable resource for collecting data for the present study, but it was not without its limitations. As noted on its own website, SCOTS is geared towards many different audiences who have myriad intentions with the data, and scholars are just one of these many audiences. As the first large-scale project of its kind in Scotland, SCOTS allows research such as the present study to be done, and its

<sup>3</sup> See footnote 1 on totals for this table.

advanced search functions made the restriction of data by certain parameters such as transcript length and genre a simple feat. From the advanced search functions, I was able to find seven recordings and 21 speakers from which to draw data. These speakers span five decades in age range; represent towns and cities in four major dialect regions; and feature in recordings from two different genres. I used Erman's 2001 monitor type framework as a way to categorize the 270 tokens of *you know* and *ye ken* found within the seven recordings, and found that her framework was adaptable to *ye ken* tokens as well as *you know* tokens; however, the line between the three monitor types was sometimes blurred, and required additional clarification in order to sort tokens into the most suitable category.

### 3 Literature review

The study of discourse markers has increased in scope and depth in the past two decades, and discourse analysis itself is a vast and ambiguous field (Schiffrin, 1987:1). *You know* is one of the more frequently studied discourse markers and is discussed in an extensive body of literature. Language scholars have investigated the function of *you know* in many different contexts, and across many sociolinguistic variables such as speaker age, gender, class, and region of residence. Much of this research has highlighted the complexity of discourse markers, and the indeterminacy of function can be seen as a hallmark of the discourse marker category (Fox Tree & Schrock, 2002:727). Investigation into the use of *ye ken* as a discourse marker is lacking, with no research to summarize.

#### 3.1 Discourse markers

Schiffrin defines discourse markers as being “sequentially dependent elements which bracket units of talk” (1987:31). Some key assumptions in the field of discourse analysis are that language always occurs in a context, is context sensitive, and is communicative. Discourse makers such as *well, but, so, I mean, and you know* do not have a single fixed semantic meaning, but rather multiple meanings that are negotiated by speakers within a given discourse (Fox Tree & Schrock 2002:727; Holmes, 1986:1; Crystal, 1988:47; Fraser, 1996:169; Erman, 2001:1339). Travis (2005:27) notes that there is little consensus in the literature about what precisely fits under the heading of “discourse marker”, and that even the term “discourse marker” is not universally used, with many other terms being used to refer to these conversational particles in addition to

this one. Among the terms mentioned in this chapter are “pragmatic expression” (Erman, 1998; Holmes, 1986); “parenthetical phrase” (Crystal, 1998); and “hedge” (Lakoff, 1975). Within this project, I will refer to *you know* as a discourse marker. This is the term most frequently used in much of the recent literature, and in addition, the term “discourse marker” does, as Travis notes, reflect the key role of such an item in a conversation, which is to mark indeterminate chunks of discourse and indicate how they should be understood in relation to context in which they occur (2005:27).

Fraser explains discourse markers as being a type of pragmatic marker. Fraser takes pragmatics to be “an account of the process by which the language user takes a sentence representation provided by the grammar and, given the context in which the sentence is uttered, determines what messages and what effects the speaker has conveyed” (1996:186). Discourse markers, then, are expressions which signal the relationship of a basic message to the foregoing discourse. They contrast with other pragmatic markers in that they only provide instructions to the addressee on how the utterance to which the discourse marker is attached is to be interpreted: they do not contribute to any representative sentence meaning (1996:186). Schiffrin notes that an analysis of discourse markers raises many questions, such as what discourse markers add to coherence; what kinds of meanings discourse markers have; and what kind of functions they have (1987:72). The answers to these questions are complex, and dependent on the specifics of a particular discourse.

### *3.1.1. You know as a discourse marker: meta-knowledge and manner of speaking*

Erman (1987:12) classifies *you know* as a pragmatic expression (PE), and she notes

that *you know* has been viewed as a marker of powerless and inferior speech, by mainstream commentators and academics alike: PEs and related expressions have been stigmatized not only by controllers of language such as teachers, but also in the academic literature. Fox Tree and Schrock make a similar observation in their summary of many earlier studies (such as Lalljee & Cook, 1975; Ragan, 1983; Fromkin, 1973:42; O'Donnell & Todd, 1991:69) which identify *you know* users as being “anxious, uncertain, or lacking in self- confidence”, and that speakers “sprinkle” *you know* and *I mean* into conversation with no distinction between the two (2002:729). Though more recent research refutes these claims, these comments are at least illustrative of how often contemporary researchers explained away the uses and functions of PEs in a way that did not closely observe the contexts in which they were uttered, and of how researchers defined in too general a way notions such as power and powerlessness.

Research into the position of *you know* within an utterance has highlighted the fact that *you know* does not have a universal function, but rather one that is influenced almost entirely by its context. Erman found in her 1987 study that *you know* occurred between arguments or propositions within a turn, and it was frequently used in a medial position as a repair marker, and also as "a staller in order to allow the speaker to do word search" (1987:51—52). *You know* in initial position was, in Erman's research, found to be a topic-shifter and turn-taking device, often with the same single token functioning in both capacities. *You know* utterance-finally did not seem to be primarily used to serve a turn-yielding function, given, as Erman notes, the low percentage of tokens occurring in this position (1987:53). Erman notes that overall, *you know* most frequently signaled hesitation, but it was also used in the organization of information into units (1987:140).

Crystal, in referencing Erman's above study in his own article, also notes that parenthetical phrases (the term he uses for *you know*) are far more complex and important than we often allow. He notes their function as giving the speaker “an opportunity to check back, to plan ahead, and to obtain listener reaction. They give the listener an opportunity to keep up and to react” (1988:48). As Fox Tree and Schrock describe, “*you know* may occur turn-initially, turn-medially, or turn-finally not because *you know* is being used to accomplish the turn coordination, but because at any point in the turn speakers may want to invite addressee references” (2002:738).

Schiffrin suggests that *you know* is a marker of meta-knowledge about what the speaker and hearer share, and also a marker of meta-knowledge about what information is generally known (1987:268). *You know* can be used to reach a situation in which a speaker becomes aware of knowledge that is also shared by the hearer, such as in the following example:

- (8) a. Jack: And when you're a cripple, you're a prej—  
b. in other words...they're cripples because they're so  
religious is what—is the point I'm trying to make.  
c. In other words they're *sick*. Religiously.  
d. Like the:... **y'know** what Hasidic is?  
e. Debby: Umhmm.  
f. Jack: The Hasidic Jew is a cripple in my eyes, a mental cripple.  
(Adapted from Schiffrin, 1987:269)

Schiffrin notes that in (d), Jack uses “*y'know* what Hasidic is?” to check that Debby knows what Hasidic means, and that in (e), Debby has the opportunity to confirm that she does indeed know what Hasidic is. In (8) overall, we see a transition from a speaker not knowing whether a hearer has knowledge of a given situation to a situation in which shared knowledge becomes apparent, and *y'know* facilitates this transition from one

information state to another. Similarly, *y'know* can be used to transition from a state in which a speaker is not aware of a hearer's *lack* of knowledge, to one in which he or she becomes aware of this lack of knowledge. There are four knowledge situations in which a speaker and hearer can find themselves, and each reflecting a different combination of what a speaker knows about a hearer's knowledge, and what a hearer actually knows (see Table 3). The conversation in (8) reflects a transition from situation (b) to situation (a).

TABLE 3. Meta-knowledge about speaker/hearer shared knowledge.

		Does speaker know of hearer's knowledge?	
		Yes	No
Does hearer know of X?	Yes	(a)	(b)
	No	(c)	(d)

(From Schifffrin, 1987:268)

Fraser (1996:181) classifies *you know* as a manner-of-speaking marker, a category within the group of commentary pragmatic markers, which are themselves “lexical expressions which have both a representational meaning specifying an entire message, and a procedural meaning signaling that this message is to function as a comment on some aspect of the basic message” (1996:179). Fraser identifies seven categories of commentary markers, including assessment markers, evidential markers, and manner-of-speaking markers. Examples of assessment markers can be seen in (9):

- (9)
- a. **Amazingly**, we got lost.
  - b. **Fortunately**, he has insurance.
  - c. **Sadly**, Mary arrived too late to meet the deadline.

Assessment markers signal the speaker's evaluation of the state of the world that is represented in the proposition. Manner-of-speaking markers, such as *you know*, are markers that speakers can use to signal comment on the manner in which the basic

message is being conveyed (1996:181). Examples can be seen in (10):

(10)

- a. **To be honest**, you need to stop now.
- b. **Bluntly**, how are you going to get him off the hook?
- c. **You know**, I think you're wrong.

In addition to the basic message that the person in (8a) needs to stop now, “to be honest” as used in this context is informing the addressee that the message is being conveyed in an honest way (1996:181).

Fraser identifies *you know* in (10c) as a pragmatic marker, which he notes is different from Schiffrin’s classification of *you know* as a discourse marker: Fraser asserts that *you know* has an interpretation analogous with “confidentially” in a context such as (10c). This provides an example of the discussion from 3.1 that there is debate on how markers such as *you know*, which are high frequency and appear in a plethora of contexts, can be classified, and Fraser has provided an example here of *you know* that he believes is more suitably classified as a pragmatic marker of subjectivity, instead of as discourse marker checking for common ground. For Fraser, a discourse marker is a sub-category of pragmatic marker (1996:186), and in contrast to pragmatic markers, discourse markers “do not contribute to the representative sentence meaning, but only to the procedural meaning”. Fraser notes that discourse markers provide instructions to the addressee on how an utterance to which the discourse marker is attached should be interpreted: a quite subtle difference from the definition of manner-of-speaking marker given above.

### 3.1.2 *You know and domains of functioning*

Britt Erman, like Fraser, classifies *you know* as a type of pragmatic marker. She considers the two main functions of pragmatic markers to be those of monitors of



discourse and of interaction, and notes that these two functions are well-established (2001:1337). In the discourse domain, pragmatic markers function as textual monitors to achieve cohesion in the discourse overall; in the social domain, the markers function as social monitors concerned with the addressee's experience of the conversation. In Erman's (2001) study, she proposes a third function of pragmatic markers: metalinguistic monitors. These are pragmatic markers with metalinguistic functions, with the job of functioning as a comment on the implications of a message's propositional content and on the speaker's intended effect with it (2001:1339). Erman's (2001) study concentrates on the pragmatic functions of *you know* in British adult and adolescent talk, noting the points at which the functions of the marker differ for each group. Below, I will discuss each of the three functional domains in more detail.

### 3.1.2.1 *Textual monitors*

Textual monitors are oriented towards the text itself and are used to indicate boundaries between topics; modes of speech (such as direct and reported speech); foregrounded and backgrounded information; and cohesive devices between sets of propositions at the textual level (Erman 2001:1342). Most centrally, they create coherence through structuring and editing the discourse. Erman identifies the most typical sub-function of *you know* at clause level as being to mark certain elements in the thematic structure: for example, a speaker who wants the listener to accept part of the information discussed as being known or given. She provides the following as an example of this, from British adolescent speech:

- (11) /.../ and we, we all buy, we all buy chips yeah, and the next minute, **you know**, we're all walking into the arcade and all these girls just come up to us and start taking chips /.../

(From Erman, 2001:1342)

Erman explains that *you know* in (11) is marking the fronted time adverbial, which she says functions here as a scene setter for subsequent events. Discourse markers functioning as textual monitors are concerned primarily with the organization of the discourse and are used to mark transitions between propositions that are largely independent, and thus the markers will mark 'moves' between arguments, states, or modes in a text (2001:1343). Erman provides the following example of this, also from British adolescent speech:

- (12) So eh **y'know** it just seems that that's how things work

(From Erman, 2001:1343)

Here, *you know* introduces the proposition *it just seems that that's how things work*. I would additionally argue that introducing the proposition with *so eh* preceding *you know* only adds emphasis to the speaker's intention to change course and introduce another topic or proposition, and thus provides even more of a marker to let the speaker know that a change of argument or state is to be anticipated. In another example, we see how textual markers can mark transitions between direct and reported speech, which as Erman notes is similar in function to quotation marks in written text:

- (13) /.../ I thought I'd warn you though. Right, he goes excuse me, why is there a bear sitting there? He goes well **you know**, don't ask okay, but just don't touch him okay, cos he's dangerous just don't, **anyway** the man gets drunk /.../

(From Erman, 2001:1344)

Erman notes that *well you know* marks the beginning of the quotation, and *anyway* marks the point at which the speaker resumes direct narration. In other cases, markers with an editing function can appear in any instance where a hesitation or repair is needed. Erman

found also that *you know* co-occurred with other markers of hesitation such as repetition, and repetition plus *like*, but that such occurrence was age-dependent (see section 2.4.1).

In examples (11) - (13) above, we can see how textual monitors are used in a variety of ways to structure a text, anchor points of relevance, and guide a listener through a portion of discourse. As Erman puts it, textual monitors "belong to the text" (2001:1340), and in a broad sense can be seen as being a part of it, though they do not contribute to its propositional content. This function is in contrast to that of a social monitor, which involves the addressee's role more directly and often affects the shape and direction of the discourse.

### 3.1.2.2 *Social monitors*

Social monitors play an important role in negotiating the meaning and management of discourse and ensuring that the channel is open between interlocutors. They also play a role in ensuring that a speaker has been properly understood, or in ascertaining whether the addressee agrees with the speaker's understanding of a certain reference in the text (Erman, 2001:1340). In this sense, they function outside the text, in contrast to textual monitors above. Erman observes that in conversation management, *you know* is used with both a turn-taking (frequently also highlighting) and turn-yielding function. Below is an example of a turn-taking and highlighting function:

- (14) <1> Er **you know** I told you about that thingy?  
      <2> What?  
      <1> (continues)

(From Erman, 2001:1346)

The speaker in a conversation, according to Erman, typically selects himself or herself, as opposed to being selected by any previous speaker (Sacks, Schegloff and

Jefferson, 1974). Another common function of the pragmatic marker in the social domain is to make sure that the listener has correctly understood the specific references made in the text, and Erman refers to this function as "comprehension-securing" (2001:1346):

- (15) <1> /.../ I hate it when you see someone being sick. They go (mimicking vomiting sound)  
<2> No.  
<1> I watched erm, **you know Warren?** He was being sick right and he was hanging over the banisters like this in these flats /.../

(From Erman, 2001:1346)

- (16) Zelda: **Y'know that eh orthopedic doctor? Y'know** that took care of Henry when he had his back problems?

Irene: Who, that Chinese doctor?

Zelda: No, the Italian. Bonzi!

Irene: Oh yeh.

Henry: He was not Italian, Zelda, he was Spanish.

(From Schiffrin, 1987:270—271)

Erman suggests that what seems to determine the occurrence of the comprehension-securing function of *you know* is the relationship between the speakers, the type of discourse, and the subject matter being discussed: a clue that genre plays a role in determining the pragmatic function of *you know* alongside relationship between participants and participant age.

### 3.1.2.3 *Metalinguistic monitors*

The two functions of *you know* described thus far in this sub-section are, as Erman highlights, already well-established in the literature. She proposes a third function, employed whenever there is a need in conversation for a speaker to check whether an addressee shares a speaker's understanding of a code, and provides as an example of this the question "Do you know what I mean?" This metalinguistic function is in use whenever the speaker underscores the illocutionary force of the utterance as a whole

(2001:1347). Markers used in the metalinguistic domain focus not on the text or the participants, but on the whole message (2001:1347). In its most frequently used form, pragmatic markers in the metalinguistic domain are concerned with illustrating emphasis, as in (17) below from adolescent British speech.

(17) <1> I didn't realize what I was doing. I dunno.

<2> You're so stupid! **You know.**

<1> Yeah, yeah erm.

Melanie was talk .... I was talking to Melanie about <unclear> and Melanie goes to me, <unclear> if you go out with him, and you realize that I'm mad, and you pull your eyes out of your head!

<2> (laugh)

(From Erman, 2001:1347)

The *you know* in line 2 is urging the listener to appreciate the force of the utterance as a whole (Erman, 2001:1347). The other way in which *you know* is used metalinguistically is as an approximator, where the speaker gives the listener "a rough but sufficiently exact idea about a certain state of affairs for the general purpose of the conversation (Erman, 1995:144). Usually such functions are fulfilled by markers such as *and so on*, *and all that*, and *or something*, but Erman did find one example of *you know* fulfilling this purpose, though not without support:

(18) She said you're, you're nice, you're pretty, **you know whatever.**

(From Erman, 2001:1348)

The more common way in which approximators are used within Erman's data is with an appealing function: that is, with the speaker appealing to shared knowledge of the world, general truths, or otherwise 'uncontroversial issues' (Erman 2001:1348). Erman notes that *you know* is a good candidate for this function in view of its lexical meaning, and provides the following examples:

- (19) Henry: I'm not a -we're all not perfect, **y'know**  
I'm not perfect Zelda, after all.

(from Schiffrin, 1987:276)

- (20) <1> Stop the tape. Don't even tape this. <2> /.../ I don't know if I'll be allowed, I'll  
have to ask cos, **you know how it is**.

(From Erman, 2001:1348)

Erman observes that in this function, speakers are quite effectively using *you know* to meet Grice's cooperative principle as a basis for communication. It is, as she notes, highly unlikely that a hearer will not accept these appeals, as they do not constitute a face threat (2001:1348).

Erman's (2001) study presents three distinct functions of *you know*, and explains them in such a way that they can be identified in many different types of discourse. I am using Erman's framework of three functions of *you know* as a way to categorize tokens of *you know* and *ye ken* in the present study and to address my first research question: "Does *ye ken* have the same functions, and range of function, as *you know* in spontaneous discourse?" Through having a framework in which I can categorize tokens of *you know* and *ye ken* by function, I can address the question of whether *you know* and *ye ken* have the same functions, and range of functions, in spontaneous discourse. The question, of course, will be how suitable this framework proves to be for analyzing the function and use of the dialect form.

### 3.2 *The role of audience: age, gender, and relationship between speakers*

While the function of *you know* is contextually determined, the appearance of *you know* is also influenced by demographic factors such as the speaker or listener's age and gender, and also by the type of conversation that is being analyzed. In this section, I will

look more closely at the data from Erman's 2001 study to demonstrate how *you know* functions differently across age and culture groups, and how this influences its appearance or non-appearance. I will also summarize research by others who have examined the role(s) of sociolinguistic variables on the use of discourse markers.

### 3.2.1 Age

Erman's aforementioned 2001 study focused on *you know* as it is used by two groups: adolescent speakers of British English (from the Corpus of London Teenager Language (COLT) corpus, recorded in 1993), and adult speakers of British English (from the London-Lund Corpus (LLC), recorded between 1960 - 1975). She also makes reference to a corpus of conversation data from middle-aged first generation Americans of Jewish origin in lower middle class Philadelphia (recorded in the 1970s and documented by Schiffrin, 1987), though she does not use any data from this corpus in her investigation into monitor types. Erman tabulated the number of tokens from LLC and COLT that fit into the category of textual, social, and metalinguistic domains respectively, with the results evident in Table 4.

TABLE 4. *You know* across three functional domains in LLC and COLT.<sup>4</sup>

	Textual		Social		Metalinguistic	
	N	%	N	%	N	%
<b>LLC</b>	228/279	81.7	43/279	15.4	9/279	3.2
<b>Colt</b>	140/282	49.6	76/282	26.9	48/282	17.0

(From Erman, 2001:1349)

The order of most to least frequently used pragmatic marker type remains the same across both corpora, but as Erman notes, there are noticeable differences between the way the

<sup>4</sup> Erman does not tabulate results from Schiffrin's data.

speakers in the two corpora use *you know* (2001:1349). The adult speakers, she says, are most concerned with using the marker in order to create a coherent text, which facilitates decoding of propositional content, while young speakers use *you know* in nearly half of all instances for other reasons (however, see Section 3.2.2. for a summary of Erman's position on the limited role that age may play).

In COLT, *you know* linking propositions in argumentative discourse is rare, but in both LLC and the American corpus, the two adult corpora, it is quite frequent<sup>5</sup>. Erman notes that it is obvious that adult and adolescent speakers are involved in different discourse types: young people seem more preoccupied with telling stories and reporting events than they are with engaging in narrative discourse (2001:1343). Therefore, Erman says, the most common context and function of *you know* as a discourse marker in the adolescent corpus is that of linking propositions in narrative and descriptive discourse. British adults use *you know* more often to elicit a response from a listener, whereas in British adolescent speech *you know* more often had a turn-taking and highlighting function (2001:1345). The use of *you know* as an emphatic function 45 times in the COLT corpus, but never in the LLC corpus, also suggests that this function of *you know* may be age stratified.

### 3.2.2 Relationship between speakers

Erman also acknowledges the role of speaker relationship in her results. When looking at social monitors, Erman found only one example used of *you know* with a comprehension-securing function in the LLC corpus, but its use was much more plentiful

---

<sup>5</sup> As noted above, Erman does make reference to Schiffrin's 1987 American corpus data in her discussion of monitor types, but she does not tabulate any of this data.



in the COLT corpus (1 token and 23 tokens respectively). She notes that the use of *you know* as comprehension securing in the Schiffrin data was not at all uncommon either. This leads Erman to propose that it is the relationship between speakers, and not their age, that is a decisive factor in the use or non-use of *you know* with a comprehension-securing function (Erman 2001:1346). She notes that in the American corpus, participants are close-knit, and this is also the case in the COLT corpus, and as such, a close-knit relationship could influence not only discourse structure but the topics discussed in the first place (see section 3.5). Fox Tree and Schrock note that people from a certain community might use *you know* more often not to show that they are members of a community, but because they are more willing to appeal to addressee inferences (2002:737). *You know* may be more common amongst friends than strangers because friends share more mutual knowledge, and “speakers may be more likely to invite addressee inferences when they believe inferences drawn will approximate their thinking” (2002:737).

### 3.2.3 Gender

The notion of gendered speech has been active for centuries (e.g. Bergvall & Bing, 1998:495), and there has long been a perception that hedges in particular are used more by women and are characteristic of ‘women’s speech’ (Lakoff 1975:54; Coates 1993:116; Holmes 1986:12; Erman 1987:26). This assertion was brought to light and given further impetus in part by work by Lakoff (1975), who made the association between hedges and women’s speech, and between women’s speech and powerlessness. Lakoff proposed that women’s speech firstly contains more hedges, and secondly and subsequently that this is

because women are “socialized to believe that asserting themselves strongly isn’t nice or ladylike, or even feminine” (Lakoff, 1975:54). Erman (1987:26), in a discussion of Lakoff’s 1975 proposal, states “the fact that we label a certain linguistic item as inferior or powerless is of course not due to inherent features of this item but is instead explained by the way we associate certain linguistic items with certain types of users.” Such critiques of gendered language use became more frequent following Lakoff’s 1975 work, with Edelsky (1993:192) critical of the way in which research into gendered language, including her own, is conducted and presented, and Holmes (1998a:479) noting that “it is important to recognize that gender-differentiated patterns of linguistic behavior will reflect the particular interaction of biological, social, and psychological factors in any society”.

Using data based on her own research, Holmes found that, in comparing *you know* in formal versus informal interactions, there was no significant difference in the overall distribution of *you know* in the speech of men and women (1986:12). Holmes additionally found that that men use *you know* more often than women to express uncertainty, whereas women use *you know* more often than men to convey confidence (Coates, 1993:117, citing Holmes 1987:64). In her study of *you know* and *I know* among adolescents, Irwin (2006:517) finds that in her research, *you know* is a feature “mostly associated with male speakers”. Holmes notes that her results differ from results in work by Fishman (1978, 1980) and Östman (1981) for American women and men, which supports Lakoff’s claim that women are more likely to use *you know* more frequently than men. However, in these latter studies, Holmes points out that the amount of speech produced by men and women in Östman’s study was unequal, with women contributing

over 60 percent of the total talk (1986:13).

One positive outcome from Lakoff's early work on language and gender is that subsequent work tended to focus more on empirical study than on generalized assertions. However, in addition to Edelsky's critique above, Erman notes "not until the contexts and functions of PEs have been meticulously analyzed in a multitude of diverse material, can we include sociolinguistic variables (e.g. education and sex) in our analysis. If we start by finding out to what extent, for instance, men and women use PEs [discourse markers] without first having established their characteristic contexts and functions regardless of the sex variable, we risk falling into ad-hoc, let alone stereotyped, explanations" (1987:217).

Though including gender as a variable in language research is complicated by the performative nature of human communication, the complexity of exploring gender's role in language should not rule out its inclusion as a variable in a study such as the present one. As such, I include gender as a variable in the present study cautiously but purposely.

### 3.3 *Genre and register*

As noted in 3.1 above, *you know* does not have a universal function but is rather influenced by context, and the work summarized in this chapter suggests that a significant factor influencing the functions and use of *you know* is the genre or register in which it appears.

Biber and Barbieri's (2007) study into lexical bundles in university spoken and written registers revealed that lexical bundles are common in all spoken university registers, and are relatively rare in written academic registers (2007:281). They define

lexical bundles for the purpose of their study as being simply “the most frequently recurring sequences of words” within a sub-corpus of texts from a single register (2007:264, 267). Though *you know* is not featured as a lexical bundle within this study, Erman (2001:1353) identifies adolescent speech as having a tendency to “come out in more or less ready-made chunks” and that “*you know* is one of those frequent combinations of words which [... forms] part of larger prefabricated structures”, and thus we can see *you know* as related in form and perhaps function to the lexical bundles in Biber and Barbieri’s study.

Erman’s 2001 study corroborates the relevance of genre to the functions of *you know*, saying that the content and organization of her three corpora affected the results of her functional domains study. There are many features in the COLT which suggest speakers are quite intimate with each other, and that as with the Schiffrin corpus, the dominant discourse type is narrative: many of the conversations among participants from Schiffrin’s 1987 corpus involve recalling past experiences in life (2001:1346). In the LLC, the characteristic discourse type is informative, descriptive and argumentative, with interlocutors frequently exchanging information and describing phenomena at work and at home, as well as putting forward ideas and beliefs (2001:1351). The discourse in the COLT is, on the whole, more “jocular and expressive”, dominated by the telling of stories and jokes, and at times quoting dialogue extensively from others. Erman states that the speakers in COLT are “always seeking to take their listener/s by surprise”, and in fact the main purpose of engaging in conversation “seems to be for the speaker to make the others laugh” (2001:1352-3). Erman concludes, based on this, that genre is a more important factor in explaining discrepancies among numbers of comprehension-securing

instances of *you know* between the corpora than age (2001:1347).

Fox Tree and Schrock (2002:737) cite Holmes (1986) and Östman (1981) in stating that *you know* is more common in narrative sections of a conversation than in sections with frequent turns, though at the same time, it is more common in opinionated talk than in narratives, although opinionated talk seems more likely to have more turns (2002:730). *You know* may be more common in opinionated talk, they propose, because speakers may desire addressees' filling out ideas along the lines of speakers' opinions (2002:737). Although results differ from study to study, as do emphases on what the exact effects of genre differences may be, it is apparent that genre is an interesting and fruitful factor to include in the study of discourse.

### 3.4 *Summary*

There is no shortage of research into the topic of discourse markers and their functions within conversations, nor into the factors that may influence the use of these discourse markers. *You know* performs several important functions in conversation management such as topic shift or maintenance; turn-yielding and turn-taking; the turn structure and content. The function of *you know* is influenced by genre; relationships between speakers; and also a speaker's age. However, all of the literature cited within this section refers to the Standard English discourse marker *you know*, and there is simply no way of knowing how *ye ken* functions in conversation, and what similarities, if any, exist between these two forms of this discourse marker.

### 3.5 Research questions

This project aims to address three central research questions:

- (1) What differences exist in the ways that *you know* and *ye ken* are used by certain groups of speakers?
- (2) What functions, and range of function, do *you know* and *ye ken* have in spontaneous discourse?
- (3) Are there any other notable differences between the ways in which *you know* and *ye ken* are used in spontaneous discourse in present day Scotland?

As mentioned in 3.1.2.3., I will be using Erman's functional domains framework to categorize *you know* and *ye ken* as functioning as textual monitors, sociolinguistic monitors, or metalinguistic monitors. Question (1) speaks to the roles that sociolinguistic variables, specifically gender, age, and geographic region, play in the selection of one linguistic form over another. For certain groups of speakers, *ye ken* may have a specific function or set of functions that *you know* does not fulfill, especially for speakers who use both varieties. Erman's 2001 study suggests that some functions of *you know* may be age stratified, and it is my prediction that the distribution of *you know* and *ye ken* will also be age stratified, with the use of *ye ken* occurring primarily in speakers born before or during World War II, and *you know* in younger speakers. Fox Tree and Schrock's claim that a certain community might use *you know* more often not to demonstrate community membership but because the group shares more mutual knowledge (2002:737) may be relevant here. If *you know* is used in this way, perhaps *ye ken* is used in both ways: to invite addressee inferences amongst a group who shared common knowledge, but also as a way to indicate community membership, and to signal that *ye ken* speakers form their

own group, be it a group of older speakers (as I suspect), or speakers from a particular region, or sharing a particular hobby.

Question (2) speaks to the roles that the interactional variables of genre and functional domains might play in the selection of one linguistic form (*you know*) over another (*ye ken*). In the present study, *you know* tokens function as a control group, with the real focus on the functions and range of functions of *ye ken*. It may be that *you know* is used in a different way than shown in Erman's study when used by speakers who also use *ye ken*, or when used in conversation with other speakers who use *ye ken*. Given that social and textual monitor functions are so commonly attested in the literature, I hypothesize that *ye ken* will perform these functions most often as well; however, it may be that speakers who use both *you know* and *ye ken* use them in complementary ways.

Question (3) allows for the fact that, given the lack of previous research into the functions of *ye ken* as a discourse marker, *ye ken* may have a subset of functions quite different from those of *you know*, especially when both *you know* and *ye ken* are used by the same speaker. The findings for question (3) can be used as a springboard for further research into the use of this discourse marker.

## 4 Methodology and Data

This chapter examines and explains the methodological choices made, and issues encountered, during the data collection and analysis for this project. No project is without its methodological issues, and in coding and analyzing my data I experienced some of the tribulations that are encountered by many researchers. Using a corpus that is web-based and also relatively recently compiled had advantages and drawbacks, both of which will be discussed here.

### 4.1 *Obtaining tokens*

#### 4.1.1 *Narrowing the data and selecting tokens*

For the purpose of studying discourse-in-interaction, I was only interested in obtaining tokens of *ye ken/you know* from spoken texts that featured a minimum of two speakers. When searching the corpus for spoken tokens of *you know* and *ye know* (for a discussion, see section 4.3.2.2.), I found a total of 4,829 tokens, and in searching for spoken tokens of *ye ken* and *you ken* (see same section), I found 171 tokens. I further narrowed this group of tokens by selecting for tokens that appeared within recordings that were made with the participants' awareness. The quality of surreptitiously recorded conversations is often lower than those made with participant awareness, as microphones are concealed and white noise and background noise are more likely. Recordings made with participant awareness also enable researches to use data in an ethical way, knowing that if a conversation participant should come across academic research into their own language use, the discovery should not be an unpleasant surprise. Refining the search in this way did not reduce the useable number of *ye ken* tokens, but did reduce the number



of useable *you know* tokens to 4,007. I further narrowed the number of useable tokens by searching only for tokens which occurred in spontaneous discourse, the discourse mode of choice for many discourse analysis studies (Erman 1987; Erman 2001; Holmes 1986; Scheibman 2002), as it enables us to study talk as it unfolds and the mechanisms involved with processing and producing language on-line.

This further narrowing of tokens produced a total of 3,906 useable tokens of *you know*, and 170 tokens of *ye ken*: an adequate size from which to draw a smaller group of tokens based on further specification. I selected spoken, spontaneous recordings of texts that were between 5,000 - 15,000 words and which contained a minimum of five *ye ken* tokens, and I selected texts that featured both *you know* and *ye ken* tokens where possible. This resulted in the selection of seven transcripts (summarized in Table 4).

#### 4.1.2 *Transcripts selected*

From the recordings of spoken, spontaneous discourse that were 5,000 – 15,000 words in length, I then further selected for texts that could be classified as belonging to one of two genres: conversational, and narrative. I also made sure to represent each dialect region in my selections, though no recordings that fit into my criteria were available for the Gáidhealtachd region. I found five texts featuring both *you know* and *ye ken* tokens and a further two that featured just *you know* tokens, and an overview of the conversations selected is as follows. From Table 5, we can see that there were a total of 270 tokens used for this study from 60,185 words of conversation over seven different transcribed recordings. It is worth noting that although the seventh recording is identified

as “BBC Voices: Aberdeen”, the participants in this recording, with the exception of the interviewer, all reside in within a small suburb of Aberdeen called Torry.

After selecting the seven texts based on the above criteria, I listened to the transcripts accompanying each of the texts, and performed a concordance search to find each of the tokens of *you know* and *ye ken* in each text. The figure of 270 tokens reflects the final number of tokens used from across the seven transcripts once 8 tokens of *you know* and 9 tokens of *ye ken* were excluded on the grounds of being more typical of canonical uses of *you know* or *ye ken* than of discourse marker uses. Examples of the canonical tokens excluded are provided below:

- (21) F646: Did **you know** this is Friday the thirteenth?  
F632: Yes I did.  
(351: Conversation about childhood memories)
- (22) M1042: Eh the other word for somebody you’ve forgotten is, “**You ken //ye ken, ye ken**” //  
F1041: //”**Ye ken** fine wha I mean” //  
M1042: //Ye just keep sayin, “**Ye ken**” an that’s it [laugh] // //  
(1448: BBC Voices Aberdeen)

The example provided in (21) shows *you know* being used to form an interrogative “did you know” followed by the complement “this is Friday the thirteenth”, to which F632 responds “yes I did”. It would be hard to argue here that *you know* has any kind of function beyond forming part of an interrogative construction, and F632’s response is an almost textbook response to an interrogative. There is no minimal response from the listener, and no overlapping: this is more of a typical question-answer format.

All tokens excluded from the present study for being canonical uses of *you know* or *ye ken* were excluded for being present within a context very similar to that present in (21), with the exception of the five tokens of *ye ken* in (22). The example in (22)

accounted for more than half of the number of *ye ken* tokens excluded, and is a rather unique example of metalinguistic awareness: metalinguistic, in this case, referring to a person's ability to reflect on the use of language, something which the speakers who are featured in this corpus are quite adept at doing (see example 31 in 5.2). M1042 provides an example of *ye ken* repeated multiple times to stand for the name of a person whose name has been forgotten, and F1041 in the very next line provides another *ye ken* idiom with “*ye ken fine wha I mean*”, perhaps equivalent to *you know very well what I mean* in SSE. Both of these examples in (22) were provided by speakers as examples of how *ye ken* can be used in various Scots dialects, and neither of them were used by speakers as a way to structure an utterance, maintain rapport with an audience, or discuss a general state of affairs. As such, they were not classified as being any kind of discourse marker.

TABLE 5. Document details for the seven featured recordings.

Document Number	Recording number	Word Count	Number of participants	<i>Ye ken</i> tokens	<i>You know</i> tokens	Dialect region	Topic of recording
1	351	5,509	2	0	37	Central	Conversation about childhood memories
2	819	3,440	4	10	6	Northeastern	Conversation about family life
3	1042	5,757	4	13	11	Northern	Conversation about dialects
4	1428	6,693	4	0	28	Central	BBC Voices: Glasgow
5	1430	11,536	5	30	9	Southern	BBC Voices: Hawick
6	1431	13,714	5	5	75	Central	BBC Voices: Dunbar
7	1448	13,536	5	14	32	Northeastern	BBC Voices: Aberdeen
<b>TOTAL</b>	--	<b>60,185</b>	<b>29</b>	<b>72</b>	<b>198</b>	--	--

#### 4.1.3 *Genre within the transcripts: narratives and conversations*

Studies that factor in genre as an area of focus have provided some illuminating insights into the form and function of discourse markers. Based on the findings of these studies, I chose to include genre as a variable in my own study, and chose a narrative genre and a conversation genre. Studies by Erman (2001), Holmes (1986), and Östman (1981) point to narratives within conversation as a fruitful site for the appearance of *you know*, and therefore I sought to find conversations in which a number of narratives within conversation were present. For my narrative genre, I selected the BBC Voices recordings. For my second conversational genre, I chose to use spontaneous conversation between friends and family members of different ages on disparate topics.

The BBC Voices recordings are a series of 300 recordings featuring over 1200 different people that were made across the United Kingdom, Channel Islands, and Isle of Man. The recordings feature a selection of dialect words and phrases elicited from local speakers by a ‘chair’ speaker, along with narratives from the local speakers about their attitudes towards language and dialect, and also their childhoods. The recordings were made between 2004 and 2005 by fifty different BBC radio journalists, one from each BBC local radio station (BBC Voices, “About Voices”). Prior to each recording, each participant to be recorded was sent a list of words for which they were asked to come up with local dialect terms. On occasions where participants reacted strongly to a particular word, conversation was encouraged about their feelings towards that particular word. Participants were also asked to talk about their own background and their shared interests. The methodology for the recordings was developed with the advice of Clive Upton, from the School of English at the University of Leeds, and it draws from the

Survey of Regional English (SuRE). The topics of the recordings are formulaic enough that they provide for some useful comparisons across texts, but their individual content is unique due to different personalities, ages, regions and attitudes. Such a set of recordings is quite useful for the present project: not only are different dialect words for the same list of concepts featured in each BBC Voices recording, but the presence of these words may, in turn, elicit other dialect words, as well as narratives based around nostalgia and shared experience: a fertile ground for the use of discourse markers to provide structure and cohesion. These recordings are a nice variation on more traditional NORM (Non-mobile, Older, Rural Male) dialect recordings, in which dialect forms were elicited in isolation from a larger context. Four BBC Voices recordings were used for the narrative genre.

The three texts from the conversation genre were rather varied in topic and structure (see Table 5 for transcript information), but they are all similar in length. In one conversation a university lecturer talks to a Glaswegian woman about her childhood; in the second a family discuss changes to their local rail station; and in the third, a family discuss local varieties of language near Auchenblae where they live. While conversation topics may vary in this group, one factor is consistent: the participants know each other well. This is also the case for the BBC Voices participants, who were often coworkers, neighbors, or all members of the same sports or hobby group (BBC Voices, “About BBC Voices”).

## *4.2 Coding tokens*

### *4.2.1 Coding for monitor type*

As mentioned in 3.3.3, I selected the monitor type framework from Erman's (2001) study as a framework for categorizing *you know* and *ye ken* tokens for the present study. Erman's framework provides a way to classify tokens into three different groups, which is a manageable number for the scope of the present study, and as discussed in section 3.3, two of the three groups (textual monitors and social monitors) are already well-established as concepts in the existing discourse analysis literature, albeit sometimes known by different names. The third category, metalinguistic monitors, is a category that I was eager to try to apply to the *ye ken* tokens, to see whether *ye ken* tokens were used with this function and to better address my second research question: "What functions, and range of function, do *you know* and *ye ken* have in spontaneous discourse?"

Coding the *you know* and *ye ken* tokens was not always a straightforward task. Many tokens did fit well into one of the three monitor type categories, but for a number of tokens, there was some difficulty selecting just one category into which the token fit best. The following are examples of tokens coded for each monitor type, with a summary of the functions of each monitor type provided in Table 6.

TABLE 6. Summary of key differences between Erman's three monitor types.

Textual Monitors	Social Monitors	Metalinguistic Monitors
<ul style="list-style-type: none"> <li>• Oriented towards the text itself</li> <li>• Indicate boundaries between topics and modes of speech</li> <li>• Function as cohesive devices between sets of propositions</li> </ul>	<ul style="list-style-type: none"> <li>• Negotiate the meaning and management of discourse</li> <li>• Ensure a speaker has been properly understood</li> <li>• Ascertain whether the addressee agrees with the speaker's understanding of a reference</li> </ul>	<ul style="list-style-type: none"> <li>• Used whenever the speaker wants to underscore the illocutionary force of the utterance as a whole</li> <li>• Illustrate emphasis</li> <li>• Used to appeal to uncontroversial general truths and shared knowledge</li> </ul>

(Information from Erman, 2001:1342—1348)

#### 4.2.1.1 Textual and metalinguistic monitors

The following tokens were coded as textual monitors for reasons given below:

- (23) M1042: We used tae say 'buggered', [laugh] **ye know?** // 'I'm buggered', **ye know?** //  
F1040: //We wouldnae have been allowed tae say that.//  
(1448: BBC Voices Aberdeen)
- (24) F826: the way they spoke was different from the way your //mates at school spoke or.//  
M903: //Wi some lads like, ye, **ye ken** like eh// well my uncle would have spoken about, "Syne we'll ging ower there", ye see //  
(1042: Conversation About Dialects)
- (25) M1015: we've been brought up tae be in a a seaside eh holiday resort, so we got  
a. an influx o holiday-makers and eh ye get a lot o these corruptions, **you know** that, not things that we would necessarily say but ye ye would hear them from  
b. Glasgow, Edinburgh, Borders, **ye know**, anywhere like that, the the things that  
c. they use so ye ye occasionally hear them. It's not uncommon but **ye know** ye locally ye wouldnae use that, 'stair rods'.  
(1431: BBC Voices Dunbar)

In (23) – (25) we see tokens of *you know* and *ye ken* that appear to “belong to the text”

(Erman 2001:1340) and work to add cohesion to a speaker’s description of an event or situation. In (23), *ye know* functions as a reported speech closer of the phrases

“buggered” and “I’m buggered” respectively, similarly to example (13) in chapter

3.1.2.1. In (24), *ye ken* appears within the utterance “like, ye, ye ken like eh// well”, as part of a hesitation on M903’s part. Erman notes that there is sometimes a co-occurrence of *you know* with *like* in a repetition or hesitation (2001:1354), and that at least 30% of the instances of *you know* that she looked at formed part of a larger structure. Here, we see *ye ken* within the structure *ye ken like eh*, illustrating that *ye ken* can also appear within a larger structure. (25) features three tokens of *you know* in a longer narrative excerpt, and in longer narratives such as this, we could expect *you know* and similar

discourse markers to appear as devices of cohesion: in (25) there is also use of “but”; “anywhere like that”; “eh”; “and eh”; and “so” alongside *you know*.

M1015’s *you know* in (25a) appears at a constituent boundary before what appears to be the start of a (subsequently abandoned) relative clause, suggesting that M1015 wanted to mark this relative clause as providing important information for his overall topic, similar to example (11) in chapter 3.1.2.1 where we see a speaker use *you know* to mark an adverbial of time. *Ye know* in (25b) could be seen as a metalinguistic monitor, if viewed as part of the larger phrase “Glasgow, Edinburgh, Borders, ye know, anything like that”, similar to Erman’s example (18) in 3.1.2.3, “She said you’re, you’re nice, you’re pretty, you know whatever”. This could fit with the idea of a metalinguistic monitor which gives the listener a “rough but sufficiently exact idea about a certain state of affairs for the general purpose of the conversation” (Erman 1995:144). However, the examples of metalinguistic monitors that Erman provides are more commentary on an assessment of a proposition, situation, or event, and not items that appear to complete a list of, in this case, locations, as the *ye know* of (25b) appears to be. Compare (25b) with (26) and (27) below:

- (26) F1041: fan ye brak awa ye see everybody was close knit, //**ye ken**, like your auntie bade roon the corner, //

F1043: //Aye.//

(1448: BBC Voices Aberdeen)

- (27) F646: Because when people, when you go to school and people jibe at you for things like that, you you you don't really, [inhale] you don't really get over it. It takes you, all your assurance disappears, //**you know**.//

F632: //mmhm//

(351: Conversation about Childhood Memories)

(26) and (27) are clearer examples of what Erman describes as a speaker checking with an addressee to see that the addressee shares the speaker’s understanding of a code: a use



of *you know* that illustrates emphasis and could be used as part the phrase “do you know what I mean?” Speakers F1043 and F632 provide affirmative responses In (26) and (27) (“aye” and “mmhm”) to F1041’s and F646’s respective appeals to shared knowledge and general truths. *Ye know* in (25b), being uttered medially within a turn, receives no such response from the addressee.

This suggests that *ye know* in (25b) is performing more of a textual cohesion function than looking for any audience response: as Erman notes, part of the purpose of the appeals that metalinguistic monitors make is that they will be accepted in some way by the addressee, but there is no addressee response to (25b). The lack of addressee response in (25b), and its listing quality rather than “general state of affairs” quality, made me classify (25b) as a textual monitor, oriented towards the text itself and its cohesion, in the same way as (25a) and (25c).

#### 4.2.1.2 Social monitors

Social monitors as described in 3.1.2.2 relate closely to Schiffrin’s 1987 idea of a marker of meta-knowledge as described in 3.1.1, as Erman describes social monitors as “ensuring that a speaker has been properly understood, or in ascertaining whether the addressee agrees with the speaker's understanding of a certain reference in the text” (2001:1340), and also as a way to regulate turns. In (28), F1054 is looking for dialect words for ‘close’:

- (28) F1054: What about a close or a  
M1042: //A closie, a closie's, a closie's different ye see I, where I was born//  
F1043: //No that's west, that's west coast again.//  
M1042: we were doon a closie, a closie's an archway,  
F1043: Mmhm. //It's got a roof.//  
M1042: //**ye know**? So, it's got a roof on it an I was born doon the closie.//

//Literally.//

F1054: //Literally, outside?//

(1448: BBC Voices Aberdeen)

M1042's turn of "we were doon a closie, a closie's an archway, ye know?" is overlapped with F1043's "it's got a roof" after she has already said "mmhm", probably in agreement with M1042's explanation of "closie". This use of *ye know* illustrates the use of a social monitor in a way that Erman summarizes as signifying "Hang on I'm not finished yet" (2001:1345). With "ye know?", M1042 regains the floor and is able to not only expand on the meaning of a closie, but also to steer the conversation into a related sub-topic.

- (29) F646: [sniff] Heard a a thing, a wee thing the day Dulcie. You know, Dulcie, you know, a girl's name?  
F632: Really?  
F646: I means "sweet".  
F632: hm  
F646: As as in, it must be, as in "dulcet".  
F632: uh-huh  
F646: **You know** how they talk about "dulcet tones", when somebody sings.  
F632: mmhm  
F646: And, I didnae know, I didnae know that; I thought it was just one of these names folk picked up, but it turns out Dulcie means "sweet".

(351: Conversation about Childhood Memories)

- (30) M822: //But then the, the// Scottish Heritage lads are nae very keen on concrete,  
M824: Well.  
M822: **ye ken** wi this bein a listed building ye see?  
F606: uh-huh

(819: Conversation about Family Life)

I coded *you know* in (29) as a social monitor because *you know* plays an important role here in helping F646 check with the addressee that she understands the idea of how a young child being called Dulcie relates to the concept of "dulcet tones". In (30), three people are discussing the interior of a railway station, with M822 saying "ye ken wi this bein a listed building ye see?" as a way to check that the audience has understood the connection between the resistance to build concrete structures at the station and the

station being a listed building, and the “ye see?” reinforces this need to check with the audience.

Social monitors were overall an easier group to code, because their role of negotiating meaning and management of discourse is a fairly common one, and because it is often clear through a token’s position within a turn when a speaker is looking to yield or retain a turn, or ensure mutual understanding of a topic. At times when a token seemed to fit quite well into two different categories (most often textual and metalinguistic, and secondly social and metalinguistic), I erred on the side of the more commonly occurring monitor types, and previous research indicating that social and textual monitors, though often referred to by other names, are two of the most frequent uses of *you know*. (26) and (27) are examples of tokens that are quite clearly focused on appeals to shared knowledge and general truths, in a way quite distinct from the other examples, and so Erman’s third category of metalinguistic monitor did prove to be a useful categorizer for a small number of tokens (see Chapter 6).

## 5 *You know, ye ken* and social groups

This section features a discussion of the results gathered from my data. I examine the roles that the sociolinguistic variables of gender, age, and region, play in speakers' selection of one linguistic form (*you know*) over another (*ye ken*), and discuss some of the implications of my findings.

### 5.1 *Speakers' results by token type*

As discussed previously, tokens of *you know* and *ye ken* were obtained from a total of 21 speakers across seven recordings. Table 7 summarizes the number and kind of tokens produced by each of the 21 speakers.

TABLE 7. Quantity and type of tokens produced by each speaker.

Speaker #	Appears in	<i>You know</i> tokens	<i>Ye ken</i> tokens	Total tokens spoken
606	819	3	0	3
646	351	37	0	37
822	819	0	10	10
823	819	3	0	3
826	1042	5	0	5
902	1042	4	7	11
903	1042	2	6	8
1004	1428	23	0	23
1006	1428	4	0	4
1010	1430	0	23	23
1011	1430	6	3	9
1012	1430	1	0	1
1013	1430	1	3	4
1014	1431	7	1	8
1015	1431	54	3	57
1016	1431	1	0	1
1040	1448	3	0	3
1041	1448	0	9	9
1042	1448	29	4	33
1043	1448	0	1	1
1054	1428, 1430, 1431	15	2	17
TOTAL		198	72	270

These 21 speakers each produced at least one token of either *you know* or *ye ken* each, with three speakers actually producing just the one token of either *you know* or *ye ken* (speakers M1012; M1016; and F1043). The average number of tokens produced per speaker is 12.8, with the median being 8 tokens. 198 tokens of *you know* feature in the present study (73.3% of all tokens), with 72 tokens of *ye ken* (26.7% of all tokens). The standard form, *you know*, is the more commonly used form. Eight of the 21 speakers produced tokens of both *you know* and *ye ken* (see discussion below).

#### 5.1.1 Use of both *you know* and *ye ken* by speakers

As noted above, eight of the 21 speakers featured in this study produced tokens of both *you know* and *ye ken*. It is not possible to know whether the 13 speakers who use only one form in the data have only one form in their repertoire, or whether it just appears as such from the limited data available. Looking more closely at the data itself does not provide any solid answers, but it does highlight the limited conclusions to be drawn exploring data through the lens of social demographics. Speakers who use both *you know* and *ye ken* appear in four different transcripts, and there are both male and female speakers using both forms, and speakers born in the 1920s, 1930s, and 1940s using both forms.

To take a closer look, within the transcript 1448 BBC Voices Aberdeen, there are four speakers who produce tokens for the present study: three women and one man, all from Torry. Two women are born in the 1930s, as well as the one man; the third woman was born in the 1930s. Speaker M1042 is the only speaker in the transcript to produce both *you know* and *ye ken* tokens, despite all speakers being resident in Torry, and two

other speakers being born in the same decade as M1042. Speaker F1041, born in the 1930s, uses only *ye ken* tokens, and speaker F1040, born in the 1920s, uses only *you know* tokens. In transcript 1430, BBC Voices Hawick, there are three male speakers out of five total speakers, and only one male speaker uses both *you know* and *ye ken* tokens. Four speakers, including the speaker who uses both forms, are from Hawick (the fifth speaker is F1054, for whom demographic information other than gender is unavailable). The interviewer for the BBC Voices transcripts, F1054, does appear in three of the aforementioned four transcripts, but she is not present in 1042, Conversation about dialects, where we see one of three women using *you know* and *ye ken* tokens, despite two women and one man being from the same region, and all four speakers being born in the same decade.

If there are speakers who only ever use one form in their linguistic repertoire, this would need to be determined by some other means. As such, it is not possible at this point to rule out the idea that all speakers featured could potentially use both forms. If it is found that some speakers do only use one form, these speakers' use of *you know* or *ye ken* alone would need to be analyzed separately.

## 5.2 Gender

TABLE 8. Use of *you know* and *ye ken* by speaker gender.

	Men		Women	
	N	%	N	%
<b>You know</b>	118	70.2	80	78.4
<b>Ye ken</b>	50	29.7	22	21.5
<b>TOTAL</b>	<b>168</b>		<b>102</b>	

As can be evidenced from Table 1, of a total of 21 speakers, 10 are men and 11 are women. Despite this almost even gender split, women, who account for 52.3% of all speakers, produce only 37.7% of all tokens, with men producing the remaining 62.2%: a difference of 24.5 percentage points in favor of men. Men do produce more tokens of both the standard form (*you know*) and the dialect form (*ye ken*), but the distribution of tokens used by the two genders is somewhat similar: 70% of all tokens spoken by men are *you know* tokens, compared to 78% of all tokens spoken by women, and almost 30% of all tokens spoken by men are *ye ken* tokens, compared to 21% by women. Men do use *ye ken* tokens more often than women do, and in looking at pure numbers, men produce more than twice as many *ye ken* tokens as women, despite there being an almost even number of men and women speakers. Men's preference for *ye ken* over *you know* echoes findings by Trudgill and myriad other sociolinguists claiming men's preference for a non-standard form; Trudgill states that "in different parts of the English-speaking world [...], female speakers have been found to use forms considered to be 'better' or more 'correct' than those used by men", and that "many societies seem to expect a higher level of adherence to social norms – better behavior – from women than they do from men" (Trudgill, 2000:72-3).

This tying in of 'better behavior' with standard language forms is nothing new, and is in fact something that the women speakers themselves discuss in data from the present study:

(31)

F1054: An what about the way ehm you were brought up tae feel about your own language and dialect? Were you ever made self-conscious about it?

F1040: //[inaudible]//

F1041: //No.//

F1043: //No.// I actually object because I I know one o my grandchildren was told tae speak proper English an I was furious. I dinnae ken far her teacher come fae or fitever. [exhale] Eh I've heard some English dialects and I find them very diff-, I suppose they would a- //think we were very difficult tae understand,//

F1040: //Mm.//

F1043: but for a teacher in a Torry school tae tell a pupil tae speak proper English is not on.

F1041: But if you ging for a job now, Sheila, if ye're bein interviewed for a job that's fit they expect ye tae dae, ye ken especially an office job, eh because they expect you tae answer phones an ye've tae answer the phones properly, proper English.

(1448: BBC Voices Aberdeen)

It is worth noting here that speaker F1043 objects strongly to the idea of her dialect being considered unacceptable in a school environment, while speaker F1041, despite saying that she was never made to feel self-conscious about her own language use, does consider “proper English” to be appropriate over the local dialect in a work environment. This illustrates that while the studies to which Trudgill refers may show that women have been found to use “more correct” forms of language, the matter is complicated by local norms alongside gender norms. Eckert (2005:4) notes that while women in general tend to be more conservative in the use of stable variables such as the reduction of *-ing*, and also of widely stigmatized forms such as negative concord, they do tend to lead in changes in progress. Also, women's preference of one form over another is not the same at all class levels, and there tends to be a greater gender differentiation higher up the socioeconomic hierarchy.



### 5.3 Speaker age

TABLE 9. Use of *you know* and *ye ken* by speaker age.<sup>6</sup>

	Born before 1940s		Born in or after 1940s	
	N	%	N	%
<b>You know</b>	83	65.3	100	79.3
<b>Ye ken</b>	44	34.6	26	20.6
<b>TOTAL</b>	<b>127</b>		<b>126</b>	

The number of tokens produced by speakers in each of the two age groups differs by just one, in favor of the group born before the 1940s. Such close totals between the two groups enable us to observe quite a clear pattern of difference in terms of which group prefers use of which form. Those born before the 1940s produce 62.8% of all *ye ken* tokens, with those born during or after the 1940s producing only 37.1% of *ye ken* tokens. Unsurprisingly, given the greater number of *you know* tokens than *ye ken* tokens in the data, both groups of speakers produce more *you know* tokens than *ye ken* tokens in absolute numbers, but the group born during or after the 1940s produces 9.3% more of the *you know* tokens. We can see from Table 9 that almost 8 out of 10 tokens used by speakers born in or after the 1940s is a *you know* token, and for speakers born before the 1940s, over one third of all tokens produced are *ye ken* tokens. The inclusion of speaker F1054's decade of birth, if known, would not change these results in any significant way: those born before the 1940s would still produce a smaller percentage of *you know* tokens and a greater percentage of *ye ken* tokens than those born during or after the 1940s.

<sup>6</sup> In calculations factoring in demographic information such as age and geographic location, the number of total tokens from which percentages were calculated was reduced to 253: 183 *you know* tokens and 70 *ye ken* tokens. This was done to account for missing demographic data for speaker F1054, who produced 15 *you know* tokens and 2 *ye ken* tokens across four recordings.

The results in Table 9 confirm many studies acknowledging that older people use older, more regionally marked forms more often than do younger people (for example Williams & Kerswill, 1999). However, as the youngest generation, those aged 25 and younger, are absent from these recordings, it would be useful to have access to recordings of these youngest speakers in order to include their data, and to see whether any younger speakers are attempting to conform to this local form common amongst older speakers, as Williams and Kerswill found with working class teenagers from Hull in Northern England, who “appear to be resisting any movement towards standardisation of accent features; the close-knit and territorially defined nature of their community appears to inhibit language change” (1999:19). This highlights a fact pointed out by Cheshire (2004), Eckert (2005), Erman (1987), Holmes (1986) and others that the social demographic categories of age, class, gender, and ethnicity are more complex and intertwined than their labels would suggest. It is too simplistic to suggest that older speakers simply use more local forms of language, and younger speakers use more standardized or innovative forms. The data here suggest that the first part of this statement is accurate, and older speakers do use more of the established local forms as compared to younger speakers, but this only presents part of the picture.

#### 5.4 Regional differences

TABLE 10. Use of *you know* and *ye ken* by dialect region.<sup>7</sup>

	Northern Scots		Northeastern Scots		Central Scots		Southern Scots	
	N	%	N	%	N	%	N	%
<b>You know</b>	11	45.8	35	59.3	129	96.9	8	21.6
<b>Ye ken</b>	13	54.1	24	40.6	4	3.0	29	78.3
<b>TOTAL</b>	<b>24</b>		<b>59</b>		<b>133</b>		<b>37</b>	

As shown in 4.4.2, there are six speakers in the present study from the Northeastern Scots region; three from the Northern Scots region; seven from the Central Scots region; and four from the Southern Scots region. However, even with a large number of speakers from this Central region, we can see that only 4 tokens of *ye ken* were produced by speakers from this region in contrast to the 129 tokens produced of *you know*, which accounts for almost 97% of all tokens produced by Central Scots speakers and which suggests a strong preference for the standard form in this region. The town of Dunbar, one of the three towns in the Central region from which tokens appear, lies 28 miles from the capital city of Edinburgh with a population of 470,000, and 28 miles from the English border. The town of Chryston is seven miles east of Glasgow, Scotland's largest city with a metro area population of 2,500,000 and a city population of nearly 600,000. The proximity of both of these towns to major cities could go some way towards explaining why *ye ken* is such a rarity here.

Conversely, even though the tokens spoken by Southern Scots speakers account for only 14.6% of all tokens, 29 *ye ken* tokens (41.4%) come from the Southern Scots dialect region and account for 78.3% of all tokens spoken by those in the Southern Scots

<sup>7</sup> See footnote 1 on number totals for this table.

region, suggesting that the opposite is true in this region than in the Central region in regards to the prominence of the *ye ken* form. One could question whether the proximity to the English border is somewhat responsible for this result, with speakers in this region looking to distinguish themselves linguistically from their English neighbors. Tokens of *you know* and *ye ken* appear in all dialect areas; however, the 72 *ye ken* tokens come from only 12 speakers (including F1054) out of a total of 21 speakers, whereas *you know* tokens come from 17 of the 21 speakers (also including speaker F1054). Of the remaining 58.6% of *ye ken* tokens used outside of the Southern Scots region, 52.7% come from northern dialect areas: the Northeastern Scots and Northern Scots dialect regions respectively, where *ye ken* accounts for 54.1% of all tokens spoken by Northern Scots speakers in the study, and 40.6% of all tokens spoken by Northeastern Scots speakers. The city of Aberdeen lies in the Northeastern dialect area, and with a population of over 210,000 it is the third most populous city in Scotland. The Northeastern town of Torry, from which tokens also appear, is very close to Aberdeen, and as such, we cannot assume from this data that the presence of a city alone will necessary reduce or increase the likelihood of dialect forms being produced.

Cities play a dynamic and complex role in language change, and they cannot be regarded as monoliths in the way in which they guide and shape change. Kerswill & Williams (1999:10) note that in British cities such as Milton Keynes, socially fluid populations made up of newcomers hinder the formation of strong local ties and the kind of close-knit networks which reinforce linguistic norms and inhibit language change. However, Glasgow is far from a linguistically homogeneous speech community, and unlike Milton Keynes, its residents have a distinct language variety that distinguishes

them from the rest of Scotland and the United Kingdom as a whole. Thus, we need to consider the role that social networks, speech communities, and class play in the use of local variants of language, and not simply view region of residence as an isolated factor.

### 5.5 Discussion

In addressing research question (2), “What differences exist in the ways that *you know* and *ye ken* are used by certain groups of speakers?”, the results displayed in 5.1 – 5.4 above have shown that the sociolinguistic variables of gender, age, and region, all have their role to play in the selection of one form of the *you know* discourse marker over another. Despite an almost even split between men speakers and women speakers in the present data, men produce almost two thirds of all tokens in the present data, possibly because they produce more speech overall. Men are more than twice as likely to produce a token of *ye ken* than a token of *you know*, and they also produced twice as many tokens of *ye ken* than women. Men and women who were born before the 1940s produced only one more token overall than those born in or after the 1940s, yet those born before the 1940s produced almost two thirds of all *ye ken* tokens, making *ye ken* a form that is preferred by men and by older speakers. *Ye ken* is also preferred by speakers from the Southern Scots dialect region over those from the Central Scots region, the residents of which have an overwhelming preference for *you know*. Both *you know* and *ye ken* are represented fairly robustly in the Northeastern Scots dialect area.

It is possible that *ye ken* is used to indicate community membership and to signal that *ye ken* speakers form their own group, as it is preferred by men as a group; older speakers as a group; and Northeastern speakers and Southern speakers as respective

groups. This provides some insight into the way that *ye ken* is used amongst certain groups of people, and how it contrasts with *you know*: a form preferred by younger speakers, women, and those living in the Central Scots region which is home to Scotland's two largest cities. However, the fact that these social groups were designated as being likely fruitful ones for interesting results by the present researcher, and by many others before me, is important to consider. As Eckert notes (2005:3), the virtues of this kind of categorization of people by social group are coverage and replicability for future studies, both of which "depend on the use of pre-determined social categories and fairly fleeting social contact with the speakers that represent those categories". Studies from the first wave of sociolinguistic research led to a treatment of variables as being markers of primary categories such as class and gender, and Eckert notes that there are most likely dynamics underlying these social categories that deserve attention from researchers (2005:5). Meyerhoff also cautions against researchers attempting to constitute communities of practice for the sake of analysis, and where a shared enterprise common to all members, one of the defining characteristics of a community of practice, is "explanatorily vacant" (2004:528). This point can be extended to a more general caveat against a practice by researchers of assuming the existence of communities amongst groups of speakers based on speakers' language data alone.

One issue arising from the use of corpora compiled by a third party instead of data collected personally, is that the amount of personal information that can be collected about speakers is restricted to the information which the creators of the corpora saw fit to, or were able to, collect. As has been discussed in this chapter, there is a great interaction between sociolinguistic factors such as age, gender, and region of residence in

determining who uses which forms of language and why. While SCOTS does include much demographic information about speakers, which can be correlated across groups, there is still a lot of information about speakers which can only really be determined when one is working with the speakers in a more direct and personal way. This may be one of the key limitations of using a corpus to determine the role of sociolinguistic variables in the use of local versus standard language forms: we cannot really observe the speech community in action on a day-to-day basis, but rather just in snapshots in the forms of recordings and transcripts of conversations. As Bauer (2004:104) notes, “different corpora assume different degrees of idealization about the speech community they attempt to represent”. Using at most three transcripts to represent a geographic region will not give us a representative idea of what a speech community is like, but it can scratch the surface and present some preliminary findings of what might be speech patterns in that area amongst a small number of individual speakers.

One advantage of using corpora to investigate language variation is that, as Holmes (1998b:63) notes, corpus studies can be used to go beyond the study of forms into functions and interactional choices. In the following chapter, I look at the functions of *you know* and *ye ken* as they are used within talk-in-interaction, and within the framework of monitor types provided by Erman (2001).

## 6 *You know, ye ken* and domains of functioning

As discussed in 3.3, Erman (2001) classifies *you know* as a pragmatic marker, and considers there to be two main functional domains of pragmatic markers: monitors of discourse, and monitors of interaction. In the discourse domain, pragmatic markers act as textual monitors to achieve cohesion in the discourse overall, and in the interaction/social domain, pragmatic markers function as social monitors concerned with an addressee's experience of a conversation. Erman proposes a third functional domain for pragmatic markers in her 2001 study: that of the metalinguistic monitor, which is used whenever a speaker needs to check whether an addressee shares an understanding of a code, or wants to illustrate emphasis. In addressing research question (2), "What functions, and range of function, do *you know* and *ye ken* have in spontaneous discourse?", I will use this chapter to explore the roles that the interactional variables of genre and functional domains play in speakers' selection of one linguistic form (*you know*) over another (*ye ken*).

### 6.1 *Functional domains and monitor types*

Textual monitors working in the discourse domain were by far the most common of all monitor types, for both *you know* and *ye ken* tokens: they account for 71.4% of all tokens. Just under two thirds of all *ye ken* tokens function as textual monitors, and just under three quarters of all *you know* tokens function as textual monitors. Metalinguistic tokens, which work as comments on the implication of a message's propositional content, are the least common of all monitor types, accounting for just 7.4% of all tokens.



TABLE 11. Distribution of *you know* and *ye ken* by functional domain.

	Textual		Social		Metalinguistic	
	N	%	N	%	N	%
<b>You know</b>	147	76.1	39	68.4	12	60.0
<b>Ye ken</b>	46	23.8	18	31.5	8	40.0
<b>TOTAL</b>	<b>193</b>		<b>57</b>		<b>20</b>	

In all three domains, *you know* is the more frequently used monitor type.

However, *you know* is more common as a token in some domains than others: it accounts for 76.1% of all textual monitors, but only 60% of all metalinguistic monitors, and 68.4% of all social monitors. *You know* and *ye ken* are much more evenly distributed in the metalinguistic monitor category, though this category does contain the small number of tokens overall. The general pattern of tokens functioning as textual monitors most frequently, then as social monitors and metalinguistic monitors second and third most frequently, is in line with Erman's findings from both the LLC and COLT corpora.

#### 6.1.1 *You know and ye ken as textual monitors*

Textual monitors are oriented towards the text itself and are used to indicate situations such as boundaries between topics, modes of speech, and cohesive devices between sets of propositions at the textual level. 147 tokens of *you know* are used as textual monitors in the data, and 46 tokens of *ye ken* are used as textual monitors, accounting for 71.4% of all tokens overall. Below is an example of a *ye ken* textual monitor from the data. In this excerpt, a speaker describes what a tenement apartment is, and aspects of tenement life. In (31) we see an example of speaker M1042 using *you know* as a textual monitor in a discussion about parents using swear words in their children's presence:

(32)

1509 M1042: My father never swore round the house. //He used tae swear at work, I'd  
1510 heard him swearin at work but//  
1511 F1043: //Oh sure.//  
1512 M1042: in the house no one swore. Nowadays, I //I hear, **ye know**, mums  
1513 swearing in front o their//  
1514 F1040: //Every second word.//  
1515 M1042: //kids an I just don't understand it.//  
1516 F1043: //Aye, the kids swearin back at them as well [laugh] oh dear.// //No that  
1517 was I I mean my father//  
1518 M1042: //It's stupid

(1448: BBC Voices Aberdeen)

In (32), M1042 uses *ye know* after the adverb “nowadays” and before the complement of “I hear”. Placing *ye know* within a constituent foregrounds the information following *ye know*, in this case the idea of mothers swearing in front of their children, which M1042 mentions as something towards which he has a negative attitude. His use of *ye know* highlights that this, in his opinion, is a “stupid” (line 1518) thing to do.

In (33), *ye ken* functions as a transition between direct speech and the rest of M1042’s narrative, which is a common textual monitor usage in Erman’s data as well as in my own:

(33)

197 M1042: Well, a tenement was at least three storeys high, sometimes four storeys  
198 high and no one ever locked a door, and people used tae come an go all the time.  
199 They'd walk intae your house at any time day or night an you just  
200 said, "Aye aye, foo're ye daein then?" **ye ken**, an an that was it. It was  
201 wonderful community living which is now gone.

(1448: BBC Voices Aberdeen)

M1042 uses *ye ken* as a direct speech closer for the phrase “Aye aye, foo’re ye daein then?”, a Scots greeting similar to “hi, how are you?”. Use of the dialect form *ye ken* here is not surprising, given that the speaker switched code into Scots for a greeting, and he

closes his code-switch with a dialect form. This phenomenon of the use of a certain form in one utterance priming a subsequent utterance that repeats the form of the first one is well attested in literature (for example, Bock 1986), and it seems that a similar situation may be occurring within the present data, as speakers use the dialect form *ye ken* following a code-switch into a local form of Scots several times. Aitken (1979:86) notes that “there is a general tendency to associate Scots expressions with other Scots expressions and English expressions with other English expressions”: following *ye ken*, M1042 reverts back to something more typical of SSE.

M1042 is of interest, as he is one of eight speakers who uses both *you know* and *ye ken* in conversations. M1042 uses *you know* for the same function as *ye ken* in (33), namely to mark a direct speech closer. In (34), he uses *you know* as a textual monitor in part of his explanation of the meaning of the word *rerr*:

(34)

- 1437 M1042: Rerr. R.E.R.R. 'Oh I feel rerr, rerr'.  
 1438 F1054: What's that mean Bob, how would ye use that?  
 1439 M1042: //Well you you//  
 1440 F1041: //It's a rare day.// //It's a [inaudible]//  
 1441 M1042: //It it's a rerr day, it's a fine day, that's rerr, that tastes good',// 'I'm feelin  
 1442 rerr'. 'I'm pleased', **ye know**, it's it's a word that's used frequently in Torry,  
 1443 in Fittie, in the whole of Aberdeen, rerr. R.E.R.R. not R.A.R.E. Okay?  
 1444 F1054: Ehm what about 'annoyed'?  
 1445 F1043: Fizzin.  
 1446 M1042: Mmhm.  
 1447 F1043: Really fizzin, even. [laugh] //Aye, just.//

(1448: BBC Voices Aberdeen)

M1042 uses *ye know* to mark a transition between his translation of “I’m feelin rerr” as “I’m pleased”, and his providing further explanation of the word’s meaning. While “I’m feeling rerr” could be considered a Scots phrase, “I’m pleased” is an SSE phrase and occurs immediately before *ye know*, perhaps interfering with a possible priming of *ye*

*ken*. M1042 produces four tokens of *ye ken*, all four of which appear to be primed by an immediately preceding use of Scots. Two of the four tokens occur within a section of Doric narrative, which was produced by the speaker at the request of the BBC Voices interviewer F1054: M1042's *ye ken* was essentially primed by another speaker who requested he used Doric for a short period of time. Only one of the other seven speakers who use *you know* and *ye ken*, M1015, produces all of their *ye ken* tokens immediately following a code-switch into Scots, one example of which can be seen in (35) below. However, at least three of the seven other speakers engage in this priming phenomenon to some degree, as can be seen in examples (35), (38), and (43).

Speaker M1015, born in the 1940s, also produces tokens of both *you know* and *ye ken*. In the BBC Voices Dunbar excerpt below, M1015 uses textual monitors of both *you know* and *ye ken* in the same narrative:

(35)

- 619 F1054: Ehm let's do 'baby', we'll start wi Colin cause he's the expert.  
 620 M1016: 'Baby'. [laugh] That's what I've got. It's pretty self-explanatory. 'Baby'.  
 621 M1017: 'Brat', 'bairn'. //"kid".//  
 622 M1015: //Oh.//  
 623 M1017: 'Brat' if they were bad, 'bairn' if they were guid, an 'kid' if ye get older.  
 624 M1015: I have tae say 'bairn' as well eh 'the bairn' cause I always say tae my  
 625 daughter, 'How's the bairn?', an that. But eh jocularly I say "Is the b-", "Have you  
 626 a. brought your brat wi ye?", **ye ken**, but that's it as a just ehm a jokin thing but that  
 627 b. is if it just as Kenny says if they've been bad, **you know**, it's eh it's just bein a  
 628 c. little brat, **ye know**, so it it carries on. Eh I don't know why I dinnae ca them it the  
 629 d. proper name rather than the, **ye know**, these adjectives but eh that's that's what  
 630 'bairn' is eh the the most common one.  
 631 F1054: Ye wouldnae say 'wean', would ye? //"Wean".//  
 632 M1015: //Glasgow.//  
 633 M1014: //No, no that's away the, it's away the west coast side that I would think.//  
 (1431: BBC Voices Dunbar)

In (35), the speakers are offering up dialect words for *baby*, and M1015 provides a narrative about the word *bairn* in use. As with example (33) above, the speaker code-

switches into Scots for his direct speech quote (line 625-6), and then uses *ye ken* in (35a) as a direct speech closer: a dialect form following a dialect phrase (though the dialect form for *bairn* ‘baby’ is not used here): another example of a Scots phrase priming the use of *ye ken*. In line 627, M1015 says “if they've been bad, you know, it's eh it's just bein a little brat, ye know,” with *ye know* in (35c) preceding a narrowing in focus of the topic, in this case the use of “brat” to refer not just to a child, but specifically a child who has been bad. The third instance of *ye know* in (35d) appears when the speaker is performing a word search for a descriptive phrase for the terms “brat” and “bairn” (the speaker settles on “adjectives”), showing a level of awareness of his own linguistic usages.

In these three examples, speakers use *ye ken* and *you know* for similar purposes, but a speaker’s age may influence how likely he or she is to use one form over another. It seems likely that code-switching between the two forms would occur in speakers of a certain age, perhaps the age at which SSE speakers and standardized language teaching started to appear in a community. The speakers in this study who use both *you know* and *ye ken* were born between the 1920s – 1950s, and thus many would be growing up during a time of increased population mobility.

In (36), F1041 tries to defend the use of “proper English” over a local variety of English in a working environment:

(36)

1566 F1054: An what about the way ehm you were brought up tae feel about your own  
 1567 language and dialect? Were you ever made self-conscious about it?  
 1568 F1040: //[inaudible]//  
 1569 F1041: //No.//  
 1570 F1043: //No.// I actually object because I I know one o my grandchildren was told  
 1571 tae speak proper English an I was furious. I dinnae ken far her teacher come fae or  
 1572 fitever. [exhale] Eh I’ve heard some English dialects and I find them very diff-, I  
 1573 suppose they would a- //think we were very difficult tae understand,//

1574 F1040: //Mm.//  
 1575 F1043: but for a teacher in a Torry school tae tell a pupil tae speak proper English  
 1576 is not on.  
 1577 F1041: But if you ging for a job now, Sheila, if ye're bein interviewed for a job  
 1578 that's fit they expect ye tae dae, **ye ken** especially an office job, eh because they  
 1579 expect you tae answer phones an ye've tae answer the phones properly, proper  
 1580 English.

(1448: BBC Voices Aberdeen)

In lines 1577-8, F1041 uses *ye ken* following an adverbial clause of condition in “if ye’re bein interviewed for a job that’s fit they expect ye tae dae, *ye ken* especially an office job”. *Ye ken* here is a cohesive device marking a transition from the general to the more specific: F1041 first says that she thinks “proper English” is important for “a job”, and then uses the adverb “especially” accompanied by *ye ken* to emphasize a narrowing of focus from “a job” to “an office job”. This more specific example gives F1041 a way to soften the blow of disagreeing with another interlocutor, and the use of *ye ken* instead of Standard English *I mean* here might be a way to signify to her audience that she does not consider herself, as she suggests in line 1569, to be self-conscious about her own language variety, even though she does believe its use might be better suited to the home and personal domains.

(32) – (36) above provide examples of *you know* and *ye ken* functioning as textual monitors, specifically as a closer of direct speech; as a way to foreground upcoming information; and as a cohesive device to transition from the narrow to the more specific. The most common function of *you know* textual monitors is as a cohesive device, followed by marking boundaries between topics; foregrounded and backgrounded information; direct or reported speech; and hesitation or repairs. The most common function of *ye ken* textual monitors is to mark direct or reported speech, followed by anchoring points of relevance; marking boundaries between topics; and as devices of

cohesion. They were less frequently used to mark hesitation or repairs. A greater percentage of *you know* tokens function as textual monitors overall than do *ye ken* tokens, by 74% to 63%.

### 6.1.2 You know and ye ken as social monitors

Social monitors play an important role in negotiating the meaning and management of discourse, ensuring that a speaker has been properly understood, and in ascertaining whether the addressee agrees with the speaker's understanding of a certain reference in the text. A total of 39 tokens of *you know* and 18 tokens of *ye ken* are used as social monitors in the data. In (37), M1015 discusses type of hat locally referred to as an “ugly”, which was worn for farm labor:

(37)

- 285 M1015: //We've got one in the museum that eh Mrs Law made for us// //and eh  
286 it's it's wire hoops//  
287 M1014: //Aye [inaudible]//  
288 M1015: and it's eh that eh red an white check that ye used tae see Arafat wearin,  
289 **you know** the //the ehm the ehm thingy,//  
290 M1014: //Aye.//  
291 M1015: but it it's ehm the it tapers it- it's eh wider at the front for you-you- the  
292 face so you get a bigger line of vision, and it's narrower at the the nape o the neck  
293 tae tae protect the neck from the the sun as as Kenny said.  
294 F1054: Good, Colin, words for 'ugly'?

(1431: BBC Voices Dunbar)

M1015's use of *you know* works as a social monitor here, to check that his addressees agree with his reference, which in this case is a reference to Yasser Arafat as a way to describe the pattern on the material from which an ugly was made. His fluency suffers following *you know* in line 289, and in 290 M1014 responds with “aye”, suggesting that M1015's attempt to ensure his audience understood his reference was successful.

Following M1014's "aye", M1015 goes on to explain the ugly in more detail.

*Ye ken* seems to perform the same basic functions as a social monitor that are performed by *you know* as a social monitor. In the example below, *ye ken* plays a role in turn management:

(38)

- 115 F902: An we used to spik about g- gaein to the shop for 'teabread', which was  
116 cookies an scones an //pancakes an stuff,/  
117 F606: //mm// //[cough]/  
118 F902: //but that's nae used even across, that wisnae even used across here when  
119 there was a baker's shop.// You wouldnae gae intae the shop here an ask for six  
120 teabread, the bakers wouldnae hae kent whit ye was askin for.  
121 F826: //So what would they have said?//  
122 F606: //mm//  
123 F902: //Well you had to say scones, pancakes, cookies or, **ye ken**?//  
124 F826: //Oh right, yeah.//

(1042: Conversation About Dialects)

In (38), F902 and F606 are discussing words for different kinds of bread used in the local bakery where they grew up. In line 123, F902 provides examples of these words, and her use of "ye ken?" serves as a way to ensure she has been properly understood by F826, who responds in the affirmative in line 124 with "Oh right, yeah". Unlike with metalinguistic monitors, this use of *ye ken* is not about having the listener appreciate the force of an utterance overall, but rather about the speaker using *ye ken* as a comprehension-securing function to ensure her audience will accept her recollection of an event as being correct. F826 indicates her willingness to accept F902's version of events with an "Oh right, yeah" in line 124. This comprehension-securing function is emphasized by the fact that *ye ken* appears at the end of an utterance: a question to which the audience feels compelled to respond. Also worth noting is that *ye ken* appears after a list of regional words for baked goods, perhaps in another case of priming: F902 is one of



the eight speakers who uses both *you know* and *ye ken* tokens.

In (39), *you know* is being used as a social monitor to check with the audience that the listener has been properly understood in his meaning of *glaikit* in line 53—4.

(39)

- 43 F1054: What does that mean, "glaikit wee battles"?  
44 M1004: [tut] Er, it's used to describe anybody that eh just er It's supposed to be  
45 derogatory, but it isnae really, you can, you can use it derogatory, but mostly in  
46 Glasgow it's used as a kind of joke wi people, eh, if they move slowly, if they're  
47 gettin older or eh just, I don't know.  
48 F1054: Ehm, can you say it again?  
49 M1004: Glaikit wee bachle.  
50 F1054: Bachle?  
51 M1004: Bachle.  
52 F1054: Ah. //[inaudible]//  
53 M1004: //Glaikit means, I think glaikit actually means squinty-eyed.// //Squ-, **you**  
54 **know?**//  
55 F1006: //Simple?//  
56 M1004: Aye, simple-lookin, aye.  
57 F1054: Mm.  
58 M1004: Just, aye.

(1428: BBC Voices Glasgow)

M1004's "you know?" gives the other speakers a chance to respond and participate in the negotiation of the meaning of *glaikit*, which continues for a few lines until M1004 in line 58 declares "Just, aye".

(37) – (39) above provide examples of *you know* and *ye ken* functioning as social monitors, specifically to check that the addressee agrees with the speaker's reference; to manage turns; and to check that the speaker has been properly understood. The most common function of *you know* social monitors is to check that the speaker has been properly understood, followed by as a turn-taking function; checking that the addressee agrees with a speaker's reference; and as a turn-yielding function. These are all different sub-categories of social monitors that were identified by Erman in her study. The most

common function of *ye ken* social monitors is also to check that the speaker has been properly understood, followed by as a turn-yielding device; checking that the addressee agrees with a speaker's reference; and a turn-taking function. A greater percentage of *ye ken* tokens function as social monitors overall than do *you know* tokens, by 25% to 19.7%.

### 6.1.3 You know and ye ken as metalinguistic monitors

Metalinguistic markers focus not on the text or the participants, but are used whenever the speaker wants to underscore the illocutionary force of the utterance as a whole. In its most frequently used form, metalinguistic monitors are concerned with illustrating emphasis. As mentioned earlier, metalinguistic monitors were fairly rare in the data. This is not surprising, given the primary purpose of discourse markers as being a way to provide cohesion to utterances, and to monitor an audience's experiences of a conversation. However, the role of metalinguistic monitor types cannot be ignored, and they provide some fairly interesting examples in the present study. In (40) below, we see an example of a metalinguistic token of *you know* from a conversation recording.

(40)

- 529 F646: But, eh anyway, the hair come back in again, thank God, but eh I'll never  
530 forget it. That's what I'm sayin, you should never jibe at kids, you used to hear  
531 them sayin to, if a if a kid had to wear glasses at school, you always got some  
532 smart-Alec, generally it was boys.  
533 F632: mmhm  
534 F646: Christened them "old four eyes" and things like that, **you know**, and it's  
535 it's no nice, it really isnae. Should never, should never hurt anybody's f- I think  
536 you'd be better, you'd be better to hit somebody,  
537 F632: mmhm  
538 F646: than hurt their feelins, Mean, you hit you hit somebody, you can always rub  
539 it, kiss it better. But you hurt somebody's feelings and eh, you cannot kiss it  
540 better.

F646's utterance "and things like that, *you know*" (line 534) provides an example of the overall metalinguistic tone to this part of the conversation. The phrase "and things like that, *you know*" encompasses the speaker's belief that children should not be teased, and it anchors the gist of the conversation as being about F646's assertions that teasing children who are different is unfair. The lack of structure to the conversation recordings allows for this kind of exchange: *you know* here is not a marker of shared experience of the world, as is *ye ken* in a number of the BBC Voices recordings; instead it marks shared opinions about the world that are relatively uncontroversial, and provides an example of a metalinguistic token functioning as an emphaser following an evaluative statement (Erman, 2001:1352). F632's minimal responses suggest that *you know* is not being used here as a turn-yielding device; F632's "mmhm" responses indicate that the nature of what F646 is saying is a more general and uncontroversial truth, not requiring any real input or debate. There is a forty to sixty year age difference between these two speakers, and the metalinguistic use of *you know* may serve to bridge this generation gap: F632 may not have been around in the 1920s, but she does presumably understand the age-old problem of school bullying. This allows the speakers to meet Grice's cooperation principle and maintain a rapport.

*Ye ken* is sometimes used metalinguistically in a similar way. Five of the eight metalinguistic tokens of *ye ken* occur within the same recording, a BBC Voices recording in Aberdeen consisting of members of the local community council, who all know each other fairly well. The recording consists mainly of discussions about childhood and World War II memories, which results in uses of *ye ken* to share memories and past

history:

(41)

397 F1041: //folk were flittin oot o Fittie intae Torry, aye an ootgo-// //an oo- an I  
398 think it was then fan ye//  
399 M1042: //Mhm.//  
400 F1041: fan ye brak awa ye see everybody was close knit, //ye ken, like your  
401 auntie bade roon the corner,//  
402 F1043: //Aye.//  
403 F1041: yer sister would hae bade across the road, //yer uncle bade at ither side but  
404 fan that//  
405 F1043: //Aye.//  
406 F1041: fan that stopped eh like my auntie bade next door an her daughter bade  
407 next door, my great grunny bade across the road, my ma bade at the ither side o  
408 the road, an we were aa sort o in a group well fan ye brak up, I think that stops.  
409 //ye nae think so? Aye, aye.//

(1448: BBC Voices Aberdeen)

In line 400, F1041 uses *ye ken* to emphasize a general truth about the former close-knit life in the area of Torry, to which F1043 responds with “Aye” (line 402). As with (40) above, the speaker who is not producing the narrative about a past experience produces just minimal responses (lines 402, 405), suggesting once again that all that is necessary here is an acknowledgement that this statement is factual: the concern is not with the audience’s input, nor is it with providing cohesion to the speaker’s narrative. In a conversation about local varieties of language in Auchenblae, two speakers also reminisce about the past in their town:

(42)

468 M903: //Oh yeah, on a Saturday nicht// you used to walk doon here fan I was a  
469 loon, ye used to walk doon the street here, every hoose on the road doon was  
470 playin country me- eh music, ye ken?  
471 F902: Scotch //country dancing.//  
472 F826: //mm//  
473 M903: //Sco- Scottish country music,// an every hoose was that like //but noo ye  
474 never hear that like.//

(1042: Conversation About Dialects)

In (42) we see an example of a metalinguistic token used in a similar way to that in (41) above, albeit with different speakers and in a different genre (a conversation recording). This token elicits more than just a minimal response from F902, who instead expands on M903's assertion that every house used to play music on Saturday nights when he was a boy. F902 confirms M903's version of events by explaining that the type of music was Scotch country dancing, and in doing so confirms that music was indeed playing.

(40) – (42) above provide examples of *you know* and *ye ken* functioning as metalinguistic monitors, specifically to describe a general state of affairs, and to emphasize general truths. The most common function of *you know* metalinguistic monitors is to check that the addressee understands the 'code' being discussed, followed by describing a general state of affairs, and appealing to uncontroversial truths. The most common function of *ye ken* metalinguistic monitors is appealing to uncontroversial truths, followed by checking that a speaker understands the code. A greater percentage of *ye ken* tokens function as metalinguistic monitors overall than do *you know* tokens, by 11% to 6%. Obviously, the numbers for both account for a very small percentage of the data overall.

## 6.2 *Functional domains and genre*

My second research question, "What functions, and range of function, do *you know* and *ye ken* have in spontaneous discourse?", addresses the role that the interactional variables of genre and functional domain play in the use of either *you know* or *ye ken* by speakers. So far I have discussed functional domains and their role in my data, and now I turn to the role of genre and how it interacts with functional domains.

TABLE 12. Distribution of monitor types across two genres of discourse.

	<b>Textual</b>		<b>Social</b>		<b>Metalinguistic</b>	
	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>
<b>BBC Voices</b>	153	79.2	26	45.6	14	70.0
<b>Conversations</b>	40	20.7	31	54.3	6	30.0
<b>TOTAL</b>	<b>193</b>		<b>57</b>		<b>20</b>	

When we look at the distribution of monitor types by genre of discourse in Table 12, we see that there are 116 more tokens featured in the BBC Voices recordings than in the conversations recordings, with 71.4% of all tokens coming from the BBC Voices recordings and just 28.5% from the conversations recordings. Four BBC Voices recordings and three conversation recordings were used for this study; the conversation recordings comprise 14,706 words total, and the BBC Voices recordings comprise 45,749 words total. The BBC Voices recordings contain over three times as many words, so it is not unexpected to find over 2.5 times as many tokens within the BBC Voices recordings as within the conversation recordings. One interesting observation from Table 12 is that social monitors are the only monitor type for which the majority of tokens appear within the conversation genre. With no interviewer present in the conversation genre participants had to negotiate turns by themselves, and this could account for the higher number of social monitor tokens in the conversation genre, given the role of social monitors in negotiating turn management.

One of the purposes of the BBC Voices recordings was to elicit conversation from all participants so as to collect as many different dialect variants as possible, as well as to collect stories of childhood and growing up. This resulted in longer recordings than for the conversation genre, the recordings for which were conducted without external

direction and which were allowed to die out naturally, and as a result, all but one of the BBC Voices recordings used for this study have word counts over 11,500. The conversation recordings were much more free form, and were presumably as short or as long as the participants felt they should be. Not one conversation recording is over 6,000 words. Also worth noting is the fact that the BBC Voices recordings each have an average of 4 participants, whereas the free-form conversations have an average of 3 participants (the BBC Voices recordings all featured an interviewer). The differences between these two genres is reminiscent of the genre differences in Erman's 2001 study, with respect to the difference between the contents and composition of recordings from the London Lund Corpus and from the Corpus of London Teenager Language: the LLC recordings were more informative and descriptive, whereas the COLT recordings were more informal and rapport-building. I would say that LLC more closely resembles the BBC Voices recordings in this respect, while COLT resembles the free-form conversations.

In comparing numbers between my findings and Erman's findings, which are summarized in Table 13, we can see that there is a higher percentage of tokens functioning as textual monitors in both the LLC and the BBC Voices recordings (81.7% and 79.2% respectively) as compared to the COLT and the conversation recordings (49.6% and 51.9% respectively).

TABLE 13. Comparison of distribution of monitor types across two genres of discourse, from Erman (2001) and data from the present study (data from Erman 2001 provided in *Italics*).

	Textual		Social		Metalinguistic	
	N	%	N	%	N	%
<b><i>LLC</i></b>	<i>228/279</i>	<i>81.7</i>	<i>43/279</i>	<i>15.4</i>	<i>9/279</i>	<i>3.2</i>
<b>BBC Voices</b>	153/193	79.2	26/193	13.4	14/193	7.2
<b><i>COLT</i></b>	<i>140/282</i>	<i>49.6</i>	<i>76/282</i>	<i>26.9</i>	<i>48/282</i>	<i>17.0</i>
<b>Conversations</b>	40/77	51.9	31/77	40.2	6/77	7.7

There is also a lower percentage of tokens used as social monitors in both the LLC and BBC Voices recordings (15.4% and 13.4% respectively) as compared to the COLT and conversation recordings (26.9% and 40.2% respectively).

It is worth noting that Erman's 2001 study looked at differences in uses of *you know* by age (see section 5.3.1 for a discussion of this experimental design difference). The present study did not factor in age in selecting the two genres of recordings used for data. Despite this methodological difference, we see similar results across both studies for textual and social monitor use by genre. Erman notes that age only goes so far in explaining her results, and that the relationship between participants and the topics of conversation being discussed were more important in explaining discrepancies amongst the numbers of monitor types used in each corpus (2001:1346—7).

If genre is indeed the more important factor, we can consider that perhaps both *you know* and *ye ken* are just more likely to appear as textual monitors in cases where there are longer narrative sections and more informative, argumentative and descriptive discourse; similarly, *you know* and *ye ken* are just more likely to appear as social monitors in cases where there is more of a focus on rapport-building and on being



“jocular and expressive”. These results suggest that discourse functions of *you know* and *ye ken* are indeed influenced by genre, and in a similar way to the results Erman found in her 2001 study. My findings confirm the work of Holmes (1986) who states that *you know* is more common in narrative sections of a conversation than in sections with frequent turns: the former describing BBC Voices quite well, and the latter describing the conversation recordings.

In the LLC and COLT corpora, tokens function as metalinguistic monitors 3.2% of the time and 17% of the time respectively. In the BBC Voices and conversation recordings, tokens function as metalinguistic monitors 7.2% of the time and 7.7% of the time respectively: proportionally speaking, almost equally. In the present data, metalinguistic tokens are slightly more commonly found in the corpus that is more informal and rapport building, as with Erman’s data. However, the percentage difference between the results from the two corpora for this monitor type is much bigger in Erman’s data than in mine. Erman was not looking at two forms of *you know*, and perhaps by breaking these results down further by standard versus dialect form, we can attempt to account for this difference in findings across the two studies.

#### 6.2.1 You know, ye ken and genre

The most key difference between my study and Erman’s study is that Erman was, of course, not investigating variation in the use of a standard and a dialect form of *you know*. This is where a comparison between the two sets of results becomes more difficult. From the above discussion it could be argued that, standard or dialect form aside, speakers in general are more likely to use discourse markers as ways to structure their

utterances and provide cohesion to talk-in-interaction than for anything other purpose: a matter discussed extensively in Chapter 3. Given that one of the main concerns of any interlocutor is to be understood by his or her audience, it makes sense that this structuring function of *you know* and *ye ken* is the most commonly used function. Table 14 below summarizes the distribution of *you know* and *ye ken* across both genres.

TABLE 14. Distribution of *you know* and *ye ken* by functional domain across two genres of discourse.

	<b>Textual</b>		<b>Social</b>		<b>Metalinguistic</b>	
	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>
<b>BBC Voices</b>						
<b>You know</b>	115	79.8	21	80.8	8	57.1
<b>Ye ken</b>	38	24.8	5	19.2	6	42.8
<b>TOTAL</b>	<b>153</b>		<b>26</b>		<b>14</b>	
<b>Conversations</b>						
<b>You know</b>	32	80.0	18	58.0	4	66.6
<b>Ye ken</b>	8	20.0	13	41.9	2	33.3
<b>TOTAL</b>	<b>40</b>		<b>31</b>		<b>6</b>	

In the BBC Voices genre, which is composed primarily of recordings that are informative and descriptive, we are mostly likely to see the standard form, *you know*, functioning as a textual monitor, working to achieve cohesion in the overall discourse. In the conversation genre, composed primarily of recordings that are informal and rapport-building, the same is true: *you know* is most likely to function as a textual monitor, working to achieve cohesion in the overall discourse. However, the results for *ye ken* highlight an interesting difference in the patterning of monitor types that we have seen so far. In the conversation genre, *ye ken* is more frequently used as a social monitor than as a textual monitor (13 tokens versus 8 tokens). Also, *ye ken* is used slightly more frequently

a metalinguistic monitor in the BBC Voices genre than it is as a social monitor (6 tokens versus 5 tokens), in a much closer result than for the conversation genre where there are only 2 metalinguistic tokens of *ye ken* versus 13 social monitor tokens. For *ye ken* in the conversation genre, there is a departure from the *you know* pattern of textual monitors, then social monitors, then metalinguistic monitors being the most commonly used. In looking more closely at the use of metalinguistic tokens, we see that more tokens of *you know* function metalinguistically than do tokens of *ye ken*, and this is true across both genres. The difference in number and percentage between the two is rather small, however, and may just be the result of the fact that there are simply more tokens of *you know* overall.

Another factor that could account for the lower numbers of metalinguistic tokens in my study as compared to Erman's is that there was not always a clear boundary between monitor types in tokens from my data, and when faced with a token that could fit into the metalinguistic token category or a textual/social monitor category, I erred on the side of caution and coded the token as one of the two more commonly occurring monitor types. There was not always a clear boundary between metalinguistic and social monitors in my data. As discussed in section 4.3.1, I coded instances where there was an appeal to shared knowledge and general truths, followed by a minimal response of support from the audience, as metalinguistic monitors; and instances where there was an appeal to the audience for confirmation of an event or circumstance coupled with a confirmation or rejection of the specifics, as social monitors (we would not expect rejection of propositional content with a metalinguistic monitor, due to the very nature of the monitor's use to share uncontroversial information). The interesting use that I will focus

on here is the use of *ye ken* and *you know* as a marker of shared reminiscence, which is a function that has characteristics in common with both metalinguistic and social monitor use. Given that only 20 tokens of metalinguistic monitors were found in the data, it is possible to look more closely at some of these tokens, particularly *ye ken*, and how they are used in context.

### 6.3 *Metalinguistic monitors: markers of shared experience*

Below are examples of metalinguistic uses of both *you know* and *ye ken* from three recordings in two different genres. The first example is from the BBC Voices: Dunbar recording, in which there appear five tokens of *ye ken* compared to 75 of *you know*, so if speaker F1054 were trying to accommodate to the other speakers with her use of *ye ken*, one could question who she was trying to accommodate to. It is, of course, possible that F1054 as interviewer is trying to elicit the use of more dialect forms by using some herself. Other factors that make this transcript unusual are that the age range of the speakers spans those born in the 1930s to those born in the 1970s; there is only one other transcript that features speakers born in at least three different decades. The four speakers within the BBC Voices Dunbar conversation are all personal friends.

(43)

- 682 F1054: //Interestingly, I was speakin in Torry an they said they used tae ca one  
 683 a. another ye know by these elaborate family connections cause o fishin again, **ye**  
 684 **know**?//  
 685 M1015: //Mmhm.//  
 686 F1054: Oor Tootsie's, Millie's, Bessie's, Jimmy's, Grace's, Bob's, Jock's loon's  
 687 b. quine, //ye **ken**?//  
 688 M1015: //Aye.// //Aye.//  
 689 F1054: //So they they could go generations back or you know make all these  
 690 relationships, it's interestin.// Ehm what about 'friend', word for 'friend' in general,  
 691 Colin, got one there?

In (43), F1054 uses two metalinguistic markers, one *ye know* (43a) and one *ye ken* (43b). In (43a), she uses *ye know* utterance-finally, providing a way to obtain listener reaction (as suggested in Crystal 1988). M1015 affirms the propositional content of F1054's utterance by responding with the minimal response "mmhm" (line 685): F1054's statement is not a controversial one. Following this minimal response, F1054 provides an example of naming conventions in Torry (line 686—687), and for such a localized naming convention, concluded with the dialect lexemes *loon's quine* 'boy's girl', it is not surprising that F1054 uses "ye ken?" in (43b) as a way to check with the audience that this is in fact shared knowledge, and that the Scots in line 683 primes this dialect form *ye ken*. Interestingly, speaker M1015 responds with "mmhm" following "ye know?" (line 683), and "aye" following "ye ken", (line 688), suggesting more priming effects at work.

Metalinguistic uses of *ye ken* almost always appear in general summaries such as this one. As noted earlier, Aitken (1979:107) identifies *ye ken*, along with a number of other phrases, as an "overt Scotticism", which she describes as "that special diction of Scottish-tagged locutions used self-consciously by many Scottish speakers as a kind of stylistic grace and as a way of claiming membership of the in-group of Scotsmen." While it is difficult to confirm or refute a proposition this vague, the latter part of this comment concerning use of certain dialectal phrases to claim membership of an in-group seems to be pertinent to this analysis. The use of *ye ken* in line 686 is not only primed by the code-switch into a local variety of Scots, but it also works as a marker within the conversation of a point of reminiscence about how things used to be. A similar example can be seen in (41) and (42) above, repeated here:

(41)

397 F1041: //folk were flittin oot o Fittie intae Torry, aye an ootgo-// //an oo- an I  
398 think it was then fan ye//  
399 M1042: //Mmhm.//  
400 F1041: fan ye brak awa ye see everybody was close knit, //ye ken, like your  
401 auntie bade roon the corner,//  
402 F1043: //Aye.//  
403 F1041: yer sister would hae bade across the road, //yer uncle bade at ither side but  
404 fan that//  
405 F1043: //Aye.//

(1448: BBC Voices Aberdeen)

(42)

468 M903: //Oh yeah, on a Saturday nicht// you used to walk doon here fan I was a  
469 loon, ye used to walk doon the street here, every hoose on the road doon was  
470 playin country me- eh music, ye ken?  
471 F902: Scotch //country dancing.//  
472 F826: //mm//  
473 M903: //Sco- Scottish country music,// an every hoose was that like //but noo ye  
474 never hear that like.//

(1042: Conversation About Dialects)

In (40), *ye ken* is used as a marker of reminiscence to describe how everyone in Torry was once close knit and lived very close together. F1043 provides a typical minimal response to this metalinguistic monitor in line 402. We do not see code-switching at work here, but instead the use of a dialect form *ye ken* within a larger chunk of a Scots dialect: speaker F1041 only produces *ye ken* tokens within the data. In (42), *ye ken* is used to reminisce about the childhood of M903, which was filled with country music. The metalinguistic use of *ye ken* serves as a marker of reminiscence and shared experience in a way that *you know* used metalinguistically does not, possibly because *ye ken* is primarily used by older speakers. Six of the seven metalinguistic tokens of *ye ken* (including the one metalinguistic *ye ken* token used by speaker F1054) are used as markers of nostalgia and shared past experiences, such as in the example below, featuring

a *ye ken* token used by a speaker born before the 1940s:

(44)

465 F1041: //I'll tell you fit did it wi us though, // my mother flitted tae Torry in  
466 nineteen thirty-nine, an it was a fortnight before war was declared. Now, because  
467 there was a war on everybody was close, everybody was close ye didnae hae that  
468 because everybody sorta had tae fit in, because you shared somebody's lobby //fan  
469 the siren went aff, //  
470 M1042: //The siren, that's right. //  
471 a. F1041: an everybody shared, **ye ken**, I mean there wasnae o- there wasnae a lot o  
472 men gaun aboot //because they were aa awa tae war. //  
473 F1043: //Mm. //  
474 F1041: So I think the war in nineteen thirty-nine did mak incomers feel mair  
475 b. welcome **ye ken** because there was that affa closeness durin the war, wasn't there?  
476 //I think so, aye. //  
477 M1042: //Aye oh yes. //

(1448: BBC Voices Aberdeen)

It is important to note that those who were participating in the BBC Voices recordings were aware that their conversations were going to be archived and serve as representative snapshots of their area's dialect; this could have had an effect on the frequency with which *ye ken* was used as a metalinguistic marker. F1041 discusses the closeness of the Torry community during World War II, and in (44a) uses *ye ken* followed by "I mean" to emphasize the reminiscent nature and "statement of uncontroversial truths" element that comes with a metalinguistic marker; in M1042's two responses, he reaffirms F1041's memories and ideas each time, once with a minimal response and once with a slightly more elaborated "Aye oh yes". In (44b), F1041 uses *ye ken* once again, in her summary of how World War II made Torry a closer community, even asking and answering her own question "there was that affa closeness durin the war, wasn't there? //I think so, aye. //"

#### 6.4 Discussion

In addressing research question (1), “What functions, and range of function, do *you know* and *ye ken* have in spontaneous discourse?”, the results displayed in 6.1 – 6.3 have shown that the roles of genre and functional domain are quite complex and interdependent. When looking at the distribution of both *you know* and *ye ken* tokens, it is apparent that textual monitors working in the discourse domain are the most common uses of both *you know* and *ye ken*. The second most common use of both of these forms is as social monitors in the social domain, and the least common use of these forms was as metalinguistic monitors used to manage a message’s propositional content. These findings confirm Erman’s findings, and also more generally the findings of many other sociolinguists who have posited that discourse markers serve primarily to structure and give cohesion to talk-in-interaction, while also serving many other interesting, yet less commonly used, functions.

When we factor in genre, we see that there is a higher percentage of tokens functioning as textual monitors in the BBC Voices recordings, the more informative and descriptive of the two genres, than there is in the conversation recordings. Conversely, there is a higher percentage of tokens functioning as social monitors in the conversation genre than there is in the BBC Voices genre. Younger speakers are less likely to produce metalinguistic tokens than are speakers born before the 1940s, albeit by a thin margin, which suggests that *ye ken* could be a marker of age, or something less obvious such as a marker of familiarity with each other, or of all being raised in the same area and sharing the same memories.



Though there were few tokens overall of metalinguistic monitors used within the recordings, the category of metalinguistic monitors proved to be a useful one for identifying a small number of tokens appearing within the seven recordings that functioned in a different way from the textual and social monitors. The focus of Erman's 2001 study was the use of metalinguistic monitors by age, specifically amongst adolescents versus older adult speakers, and although I did not focus on age in the same way that Erman's study did, I still saw similar genre effects for the distribution of monitor types. Erman does concede that the relationship between participants was a more important factor than age in explaining her results, but this does highlight the fact that looking at sociolinguistic variables, such as those discussed in Chapter 4, can illuminate patterns found within discourse data. Similarly, a discourse analysis of talk-in-interaction can illuminate sociolinguistic data, with both approaches working together to give a more in-depth picture of what happens between interlocutors during spontaneous conversation.

## 7 Conclusion

Marshall (2002:172) notes that “Urban and rural speech communities often differ from each other in marked ways. The structure of local social networks, people’s attitudes, the direction of language change, and social class structures may differ radically, so as to have dissimilar effects on dialect maintenance”. Data for the present study, as mentioned in section 4.4, was taken from speakers who live in both rural and urban areas; have different degrees of closeness; and differing attitudes towards their own dialects and the outsiders who are arriving in their area (in the case of speakers in rural locales).

I have shown in this work that *ye ken* is more likely to be used as a discourse marker by men; older speakers; and speakers residing in the Southern Scots and Northeastern Scots regions. It is proportionally most likely to be used as a metalinguistic monitor to appeal to general truths, but in absolute numbers, most likely to be used as a textual monitor to structure and give cohesion to talk-in-interaction. It features most commonly in a genre of recordings that are primarily informative and descriptive, and which also exist to preserve recordings of regional and dialect talk in the British Isles. *You know* is more likely to be used by women, young speakers, and speakers residing in the Southern Scots and Northern Scots dialect regions. As with *ye ken*, it features most commonly in a genre of recordings that are primarily informative and descriptive. One unique function of *ye ken* in the data is its use as a marker of shared past experiences and reminiscence.

The monitor type framework presented in Erman’s 2001 was beneficial in the present study to a point. It proved a useful tool for illustrating the similarities that exist

between the functions of *you know* and *ye ken* in spontaneous discourse. Given that the vast majority of residents of present-day Scotland are exposed to SSE via the domains of education, mass media, and the workplace, we can expect that a significant number of Scottish residents perform code-switching in daily life between SSE and their own local or regional language varieties. The fact that eight of the 21 speakers whose data was used in this study use both *you know* and *ye ken* as discourse markers is a testament to this environment of linguistic flexibility, and we should never underestimate the ability of speakers to be creative and innovative in their language use. Erman's monitor type framework was useful for highlighting the fact that *you know* and *ye ken* are both used as textual, social, and metalinguistic monitors throughout the seven recordings featured. However, her framework proved less useful in highlighting the differences between the two forms, although the metalinguistic monitor category was a useful third category in which to place tokens that seemed to be functioning in a less typical way within the discourse.

This directs me towards my third research question: "Are there any other notable differences between the ways in which *you know* and *ye ken* are used in spontaneous discourse?" It makes sense that *you know* and *ye ken* might have some specificity of function amongst the eight speakers who use both forms in conversation. As shown in section 6.1, switching code to Scots phrases and utterances often primes the use of Scots *ye ken*, with the standard form *you know* being used by the same speaker as little as one turn later. This priming happened with tokens that were textual monitors, social monitors, and metalinguistic monitors. We could say that priming explains the selection of either *you know* or *ye ken*, given that speakers will turn to structures that have occurred

previously in the discourse (be they in regional varieties or standardized ones), and use these structures as partial models on which to base the morphosyntax of subsequent utterances (Travis, 2007:101). Speakers will use *you know* when speaking in a variety closer to SSE on a dialect continuum, and *ye ken* when speaking in a variety closer to Scots. The phenomenon of priming might explain why a single speaker might alternate between *you know* and *ye ken* in the same genre, when speaking with the same audience, and when discussing the same topic. This usage is not watertight, however, and the priming that does occur is much more common in the BBC Voices genre than the conversation genre. It would be interesting to include a number of additional genres in further study to see what genre effects are at work.

Section 6.3 details my finding that *ye ken*, when used as a metalinguistic monitor, frequently functions as a marker of reminiscence between older speakers who have established personal or professional ties, such as with the speakers in the BBC Voices Aberdeen recording, who all work together and grew up in the same location. Earlier I referenced Fox Tree and Schrock (2002:737) as stating that *you know* may be more commonly used between friends, because groups of friends share more mutual knowledge and because friends are more likely to appeal to addressee inferences. If *you know* is more commonly used between friends because these friends share more mutual knowledge, then the use of *ye ken* amongst speakers who have grown up together in a small community, such as Torry in Aberdeen, or Dunbar, can function as a way to mark the sharing of mutual knowledge based on past experiences, and invite addressee responses in a dialect form that the dialect speakers themselves expressed concern might disappear. These same speakers who are using *ye ken* as markers of reminiscence,

discussed in example (30) the issues that arise when teachers and others viewed as “language authorities” insist on the teaching of a standard variety of language. This highlights the role that language attitudes play in the use of certain linguistic forms over others, and the role of language attitudes in the maintenance and preservation of regional language varieties.

The role of corpora in documenting language varieties and language attitudes, as well as presenting a wealth of discourse data for analysis, cannot be understated. It is because of an increased interest in the collection of regional language data for corpus use that studies such as the present one are possible. As Bauer (2004:97) notes, data contained within corpora such as SCOTS enable us to consider mechanisms of linguistic change in a systematic way, and they have the virtue of replicability for further study. Through use of data from SCOTS, I was able to highlight some key differences in the distribution and function of *you know* and *ye ken* in talk-in-interaction. I have shown in this study that there is an interplay between the role of sociolinguistic variables and interactional variables in the selection of *you know* and *ye ken* by speakers. There is currently a dearth of studies exploring the use of regional discourse markers, and I hope that studies such as this one, aided by the availability of corpora of regional language varieties, will mark the beginnings of a reversal of this trend.

## REFERENCES

- AITKEN, A. J. 1979. Scottish speech: a historical view with special reference to the Standard English of Scotland. *Languages of Scotland*, ed. by A.J. Aitken and T. McArthur, 85—118. Edinburgh: Chambers.
- AUER, PETER. 2005. The construction of linguistic borders and the linguistic construction of borders. *Dialects across borders: selected papers from the 11th international conference on methods in dialectology (Methods XI), Joensuu, August 2002*, ed. By M. Filppula, J. Klemola, M. Palander, and E. Penttilä, 3—30. Amsterdam: John Benjamins.
- BAUER, LAURIE. 2004. Inferring variation and change from public corpora. In Chambers, Trudgill, and Schilling-Estes, 97—114.
- BBC VOICES. 2007. About Voices. Online: [http://www.bbc.co.uk/voices/yourvoice/about\\_voices.shtml](http://www.bbc.co.uk/voices/yourvoice/about_voices.shtml).
- BERGVALL, VICTORIA and JANET M. BING. 1998. The question of questions: Beyond binary thinking. *Language and gender*, ed. by Jennifer Coates, 495—510. New York: Blackwell.
- BIBER, DOUGLAS and FEDERICA BARBIERI. 2007. Lexical bundles in university spoken and written registers. *English for specific purposes* 26.263—86.
- BOCK, J. KATHRYN. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18.355—87.
- CHAFE, WALLACE. 1994. *Discourse, consciousness and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.
- CHAMBERS, J.K.; PETER TRUDGILL; and NATALIE SCHILLING-ESTES (eds.) 2004. *The handbook of language variation and change*. Oxford: Blackwell.
- CHESHIRE, JENNY. 2004. Sex and gender in variationist research. In Chambers, Trudgill, and Schilling-Estes, 423—43.

- COATES, JENNIFER. 1993. *Women, men and language*. London: Longman.
- CRYSTAL, DAVID. 1988. Another look at, well, you know... . *English Today* 4.47—49.
- DOUGLAS, FIONA. 2003. The Scottish Corpus of Texts and Speech: Problems of corpus design. *Literary and Linguistic Computing* 18(1).23—37.
- EAGLE, ANDREW. 2001. Wir ain leid: an innin tae Modern Scots. Online: <http://www.scots-online.org/grammar/index.asp>.
- ECKERT, PENELOPE. 2005. *Variation, convention, and social meaning*. Paper Presented at the Annual Meeting of the Linguistic Society of America. Oakland: CA.
- EDELSKY, CAROLE. 1993. Who's got the floor? *Gender and conversational interaction*, ed. by Deborah Tannen, 45—89. Oxford: Oxford University Press.
- ERMAN, BRITT. 1987. *Pragmatic expressions in English: a study of you know, you see, and I mean in face-to-face conversation*. Stockholm: University of Stockholm dissertation.
- ERMAN, BRITT. 1995. Grammaticalization in progress: the case of *or something*. *Papers from the XVth Scandinavian conference of linguistics*, ed. by Inger Moen, Hanne Gram Simonsen and Helge Løderup, 136—47. Oslo: University of Oslo.
- ERMAN, BRITT. 2001. Pragmatic markers revisited with a focus on *you know* in adult and adolescent talk. *Journal of Pragmatics* 33.1337—59.
- FISHMAN, PAMELA M. 1978. Interaction: The work women do. *Social Problems* 25.397—406.
- FOX TREE, JEAN E. and JOSEF C. SCHROCK. 2002. Basic meanings of *you know* and *I mean*. *Journal of Pragmatics* 34.727—47.
- FRASER, BRUCE. 1996. Pragmatic markers. *Pragmatics* 6(2).167—90.

- FROMKIN, V.A.. 1973. Introduction. *Speech errors as linguistic evidence*, ed. by V. A. Fromkin, 11—45. Paris: Mouton.
- GIDDENS, A. 1997. *Sociology* (3rd edition). Cambridge: Polity Press.
- GRANT, NIGEL. 2000. *Multicultural education in Scotland*. Edinburgh: Dunedin Academic Press.
- HOLMES, JANET. 1986. Functions of you know in women's and men's speech. *Language in Society* 15(1).1—21.
- HOLMES, JANET. 1987. Hedging, fencing and other conversational habits: an analysis of gender differences in New Zealand speech. *Women and language in Australian and New Zealand society*, ed. by Anne Pauwels, 59—79. Sydney: Australian Professional Publications.
- HOLMES, JANET. 1998a. Women's talk: the question of sociolinguistic universals. *Language and gender: a reader*, ed. by Jennifer Coates, 461—483.
- HOLMES, JANET. 1998b. Generic pronouns in the Wellington corpus of spoken New Zealand English. *Kotare* 1.32—40.
- IRWIN, ANTHEA. 2006. London adolescents (re)producing power/knowledge: *You know and I know*. *Language in Society* 35.499—528.
- LALLJEE, M.G., and J.M. COOK. 1975. Anxiety and ritualized speech. *British Journal of Psychology* 66(3). 299—306.
- LAKOFF, ROBIN. 1975. *Language and woman's place*. New York: Harper Colophon.
- MARSHALL, JONATHAN. 2000. Testing social network theory in a rural setting. *Reading working papers in linguistics*. ed. by Richard Ingham and Paul Kerswill, 123—74. Reading: University of Reading.



- MCCLURE, J. D. (1994) English in Scotland. *The Cambridge history of the English language, vol V: English in Britain and overseas, origin and development*, ed. R. Birchfield, 23—93. Cambridge: Cambridge University Press.
- MCEWAN-FUJITA, EMILY. 2011. Language revitalization discourses as metaculture: Gaelic in Scotland from the 18<sup>th</sup> to 20<sup>th</sup> centuries. *Language & Communication* 31.48—62.
- MEYERHOFF, MIRIAM. 2004. Communities of practice. In Chambers, Trudgill, and Schilling-Estes, 526—48.
- MURDOCH, STEVE. 1996. Language politics in Scotland. Aberdeen University: Scots Leid Quorum.
- ÖSTMAN, J-O. 1981. *You know: a discourse functional approach*. Amsterdam: John Benjamins.
- RAGAN, S.L. 1983. Alignment and conversational coherence. *Conversational coherence: form, structure and strategy*, ed. by R.T. Craig and K. Tracy, 157—71. Beverly Hills: Sage Publications.
- REDECKER, G. 1990. Ideational and pragmatic markers of discourse structure. *Journal of Pragmatics* 14.367—81.
- SACKS, HARVEY, SCHEGLOFF, EMANUEL, AND JEFFERSON, GAIL. 1974. A simplest systematics for the organization of turn-taking in conversation. *Language* 50(4).696—735.
- SCHIFFRIN, DEBORAH. 1987. *Discourse markers*. Cambridge: Cambridge University Press.
- SCOTTISH CORPUS OF TEXTS AND SPEECH. Online: <http://www.scottishcorpus.ac.uk>
- SMITH, JENNIFER, DURHAM, MERCEDES, and FORTUNE, LIANE. 2007. “Mam, my trousers is fa’ in doon!”: Community, caregiver, and child in the acquisition of variation in a Scottish dialect. *Language Variation and Change* 19.63—99.

- TRAVIS, CATHERINE. 2005. Discourse markers in Colombian Spanish: a study in polysemy. Berlin: Mouton de Gruyter.
- TRAVIS, CATHERINE. 2007. Genre effects on subject expression in Spanish: priming in narrative and conversation. *Language variation and change* 19. 101—35.
- TRUDGILL, PETER. 2000. *Sociolinguistics: An introduction to language and society*. London: Penguin.
- TRUDGILL, PETER AND JEAN HANNAH. 2002. *International English: A guide to the varieties of Standard English*, 4<sup>th</sup> ed. London: Arnold.
- WELSH, IRVINE. 1998. *Filth*. New York: W.W. Norton.
- WILLIAMS, ANN AND PAUL KERSWILL. 1999. Dialect leveling: change and continuing in Milton Keynes, Reading and Hull. *Urban voices: accent studies in the British Isles*, ed by Paul Foulkes and Gerard Docherty, 141—62. London: Arnold.