

Response tokens in British and Irish discourse Corpus, context and variational pragmatics¹

Anne O’Keeffe and Svenja Adolphs

*Mary Immaculate College, University of Limerick and University of
Nottingham*

1. Introduction

The aim of this paper is to look at two varieties of English, British English² and Irish English, within the framework of variational pragmatics, using corpus linguistics as a methodological tool in order to assess its usefulness (and limitations) to research in this area. Our analytical focus for this investigation is response tokens, a discourse feature seen as a core part of spoken grammar (McCarthy 2002, 2003, Carter & McCarthy 2006).

2. Corpus linguistics and language variation

2.1 What is corpus linguistics?

Aijmer & Altenberg (1991: 1) describe corpus linguistics (CL) as the study of language on the basis of text corpora. It has developed rapidly since the 1960s, largely due to the advent of computers and their increasing capacity to store and process greater amounts of data. This has facilitated the systematic analysis of vast amounts of language and, in turn, has meant that descriptions (and prescriptions) about the English language have frequently been contradicted by corpus linguists who work with representative samples of naturally-occurring language (Holmes 1988, Baynham 1991, Boxer & Pickering 1995, Kettemann 1995, Baynham 1996, Carter 1998, Hughes & McCarthy 1998, and McCarthy 1998). CL is increasingly being applied to contexts and domains where the *use* of language is the focus of empirical study in a given context. Among the many fields where CL is being adopted to complement other methodological tools, such as discourse analysis and conversation analysis, are contexts such as: courtrooms (including forensic linguistics) (see Cotterill 2004), workplace discourse (Koester 2006), classroom and educational contexts (Farr 2002, 2003, Walsh 2002, O’Keeffe & Farr 2003), political discourse (McCarthy & Carter

2002), advertising and the media (O’Keeffe 2002, 2006, Charteris-Black 2004, O’Keeffe & Breen in press) and healthcare discourse (Adolphs et al. 2004).

A *corpus* is defined as “a large and principled collection of naturalised [computerised] texts” in spoken or written form (after Biber et al. 1998: 4), which is available for analysis using corpus software packages (for further definitions see Renouf 1997, Sinclair 1997, Tognini-Bonelli 2001). Some debate exists as to whether CL is a theory or a method (see Tognini-Bonelli 2001), or indeed, whether it is a new or separate branch of linguistics. Kennedy (1998) suggests that corpus-based research derives evidence from texts and so it differs from other approaches to language which depend on introspection for evidence. In this paper, we argue that a corpus-linguistic approach benefits the analysis of a pragmatic feature across varieties.

Here, we will use spoken data as our focus. Recent years have seen a major growth in the creation and development of spoken corpora, particularly in the English language, but not exclusively (see McCarthy & O’Keeffe 2004). However, many of the spoken corpora thus collected are addenda to much larger written corpora. The British National Corpus (BNC), for example is a 100 million word corpus of which 10 million words make up the spoken component (Aston & Burnard 1998). Other large well-known corpora include the Longman Spoken American Corpus (Stern 1997) and the American National Corpus (ANC) (Ide & Macleod 2001, Ide et al. 2002), of which all contain spoken components.

A number of exclusively spoken corpora are emerging which have been designed to give greater representation to spoken discourse, for example the Cambridge and Nottingham Corpus of Discourse in English (CANCODE) (McCarthy 1998, Carter 1998), five-million words collected mainly in Britain and the Limerick Corpus of Irish English (LCIE) (Farr et al. 2002), both of which will be used in this study (see below). Other spoken corpora include: the *Santa Barbara Corpus of Spoken American English*, based on 1000 of recordings of spontaneous speech from all over the United States (Du Bois et al. 2000); the two-million word *Hong Kong Corpus of Spoken English* (HKCSE) (see Cheng & Warren 1999, 2000, 2002); the *Michigan Corpus of Academic Spoken English* (MICASE), which makes almost two million words of spoken language available online (Simpson et al. 2000); the *Corpus of Spoken Professional American English* (CSPA), a two-million word corpus made up of academic discussions, committee meetings and White House press conferences (Barlow 2000); The *International Corpus of English* (ICE) project, which comprises spoken data for the Englishes of Hong Kong (Bolton et al. 2003), New Zealand (Holmes 1996), Singapore (Ooi 1997), Great Britain (Nelson et al. 2002), Ireland (Kallen & Kirk 2001), Nigeria (Banjo 1996), and the Caribbean (Nero 2000), with others under development. The *Vienna and Oxford International Corpus of English* (VOICE), a spoken corpus of English as a Lingua Franca, is also underway (see Seidlhofer 2001). In addition, *The Louvain International Database of Spoken English Interlanguage* (LINDSEI), set up in 1995 (see De Cock 1998, 2000), provides spoken data for the analysis of learner language (see also Granger et al. 2002). These many sources of data

provide researchers of spoken language with opportunities to look at large amounts of data whereas previously they might have only looked at qualitatively much smaller amounts, such as single conversations, for example. In recent years, many differences between spoken and written lexico-grammar have been identified and quantified. Carter & McCarthy (1995), for example, in one of their early works on the CANCODE corpus, introduced the term “spoken grammar” as distinct from “written grammar”. This notion of spoken grammar has been elaborated on and codified subsequently in corpus-based grammars, such as Biber et al. (1999) and Carter & McCarthy (2006).

Here we are interested in looking at a discourse feature across two language varieties, using data from two language corpora which have been assembled with the study of spoken discourse in mind, namely the Cambridge and Nottingham Corpus of Discourse in English (CANCODE) and the Limerick Corpus of Irish English (LCIE). Both CANCODE and LCIE have been designed using the same data collection and categorisation matrix (for an extensive description of CANCODE see McCarthy 1998 and for LCIE see Farr et al. 2002). These corpora are suited to variational research as they have been designed using the same principles, which are to be sensitive to speaker relationship, context and speech genre. Their data come from distinct socio-cultural settings, Britain and Ireland respectively. In this paper we will focus on a common feature of spoken interaction, namely *listener response tokens*, in the two corpora so as to examine the degree of variation, if any, between their form and use in British and Irish English.

2.2 How to build a corpus

By way of background, we illustrate how a basic language corpus is assembled from spoken and/or written texts:

Table 1. Stages of building a corpus

Spoken	Written
1) Create a design rationale	1) Create a design rationale
2) Record data (one hour of conversation yields about 12,000 words)	2) Input texts (e.g. download, scan or type and save as text files ³)
3) Transcribe recordings and save as text files	3) Database texts (classify data according to variable such as theme, genre, author, date, source etc.)
4) Database texts (classify data according to, e.g. speaker, name, age, gender, level of education, place of birth, etc.)	
5) Check transcription	

2.3 Representativeness

The issue of how best to represent a language is a key concern to corpus designers (see Atkins et al. 1992, Biber 1993, Crowdy 1994, Tognini-Bonelli 2001, Farr et al. 2002, Adolphs 2006). In the case of variation-sensitive spoken corpora, there are two core concerns:

- 1) How to best represent a language variety
- 2) How to best represent a spoken language

The first is a question of geographical and demographic coverage and sampling. For example, the British National Corpus 10-million-word spoken component consists of unscripted informal conversation recorded by volunteers selected from different age groups, regions and social classes in a demographically balanced way (Crowdy 1994). The second concern is a more complex matter relating to how spoken language itself is represented. Because most spoken corpora started out as appendage to much larger written corpora, many of them were based around written text typologies. The Cambridge and Nottingham Corpus of Discourse in English (CANCODE), which was mainly recorded in the 1990s, is one of the few corpora that has been designed to represent both a language variety and the genres of casual conversation. Recorded in a wide range of areas across Britain and Ireland, the CANCODE corpus is carefully categorised according to the relationship that holds between the speakers and according to a broad discourse goal that the speakers pursue (see McCarthy 1998 for a detailed discussion of the design rationale for CANCODE).

In terms of speaker relationships, CANCODE is divided into five broad categories which reflect the degree of familiarity between the speakers. The relationship categories are as follows: intimate, socio-cultural, professional, transactional and pedagogic. Conversations that have been assigned to the intimate category tend to take place between members of the same family or between partners, while the socio-cultural category encompasses interactions between friends. The professional category captures discourse that is related to professional interactions. The transactional category refers to situations in which the speakers do not know one another prior to the conversation that is being recorded. Typical examples would be an interaction between a customer and a waitress at a restaurant. The pedagogic category includes interactions that take place between students and lecturers or between pupils and teachers in the given institutional context. In terms of goal types, there are three broad categories in the CANCODE corpus: *information provision*, which is characterised by uni-directional interactions; *collaborative idea*, which refers to bi-directional discourse; and *collaborative task*, which includes interactions in which the participants are engaged in a task, such as assembling flat-packed furniture, for example. Since discourse is dynamic in nature, the goal-type categories sometimes change within individual conversations. Where this happened, the inter-

action was assigned to the category which reflected the dominant goal type in the interaction.

The Limerick Corpus of Irish English has been assembled using the same design matrix as the CANCODE corpus in order to facilitate inter-varietal research between British and Irish English in the given genres. Because the CANCODE and LCIE corpora have been designed to represent spoken discourse, they are suited to our purpose of looking at the discourse feature of listener response. We will now survey the existing research into this discourse feature.

3. Response tokens

Researchers from a variety of perspectives have long recognised that conversations contain listener responses, that is short utterances and non-verbal surrogates (e.g. head nods) (see Fries 1952, Kendon 1967, Yngve 1970, Maynard 1989, 1990, 1997, Tottie 1991, Drummond & Hopper 1993a, 1993b, McCarthy 2002, Gardner 2002). These signals are produced by the listener, according to Kendon (1967), as an accompaniment to a speaker. Kendon suggests that there is some evidence that the speaker relies upon these signals for guidance as to how the message is being received. Examine, for example, how the word *yeah* functions in this extract from a radio phone-in (taken from the LCIE). Here an elderly caller to a radio phone-in is explaining how, when she was young, a local woman used to do home ear-piercing, using a thick darning needle, olive oil, some string and a cork.

- (1) [for transcription conventions for this and subsequent examples of data, please see Appendix]

Caller: The way this was done was a Scottish lady who lived across the road from us.

Presenter: *Yeah*.

Caller: And she would soak some grey wool. A length of grey wool in a saucer with olive oil.

Presenter: *Yeah*.

Caller: And then she'd thread it through an extremely large darning needle.

Presenter: *Yeah*.

Caller: Then there was a cork held together... and she just threaded the needle with the wool straight through your ear and into the cork...

[LCIE]

In extract (1) we see that the presenter wants to signal that she is listening and that she wants the caller to continue telling her story, but she does not want to take over the speaking turn (or the "floor"). To achieve this, she uses short *response tokens* that keep the conversation going (in this case, *yeah*). Tottie

(1991: 255) provides an apt metaphor for this phenomenon saying that these tokens “grease the wheels of the conversation but constitute no claim to take over the turn”.

Many terms exist for this phenomenon in the research literature, often depending on discipline and definition. Yngve (1970) introduced the term *back-channel* to refer to the “short messages” that a speaker receives while holding the floor (1970: 568) and this term is widely used by many researchers. Fellego (1995) uses the term *minimal response* which comes from the body of research into language and gender (see Zimmerman & West 1975, Fishman 1978 and Coates 1986), while in another study Roger et al. (1988) use the broader term *listener response*. In this paper, we will use the term *listener response* as an umbrella term to refer to the activity involving vocal, verbal and non-verbal non-floor-holding responses when a listener responds to the floor-holding message in a conversation. We will also refer to items which are used in this activity as *response tokens*. It is worth noting that we refer here to the discourse function of these lexical items, rather than their word-class identity as adjectives or adverbs, etc. On a discourse level, Mott & Petrie (1995) point out that response tokens are the antithesis of interruptions. Duncan & Niederehe (1974) note that they project an understanding between speaker and listener that the turn has not been yielded, but they also note that it is often difficult to identify the boundary between brief utterances and proper turns where the “listener” becomes the “speaker”. This problem, however, is more for the analyst than the actual conversational participants, who, in real-time conversation, will draw on clues, such as prosodic features, facial expressions and gestures, to interpret whether an interlocutor is trying to take the floor or display listenership in a given context.

3.1 Forms of listenership

In this study, we will compare and contrast the distribution of forms and functions of such listener response tokens in two varieties of spoken English, British and Irish, using data from two corpora, CANCODE and LCIE, which have been designed for the study of spoken discourse, both using the same design matrix as detailed above. The existing research on forms shows that response tokens can be divided into *minimal* and *non-minimal* response tokens (Fellego 1995, Zimmerman & West 1975, Fishman 1978, Schegloff 1982, Maynard 1989, 1990, 1997, Gardner 1997, 1998, 2002, Tottie 1991, McCarthy & Carter 2000, McCarthy 2002). Usually, minimal responses are defined as short utterances (for example *yeah*) or non-word vocalisations (such as *mm*, *umhum*), while non-minimal response tokens are mostly adverbs or adjectives functioning as pragmatic markers (e.g. *good*, *really great*, *absolutely*) or short phrases/minimal clauses (e.g. *you’re not serious*, *Is that so?* *by all means*, *fair enough*, *that’s true*, *not at all*). The distinction is, however, not necessarily clear cut, especially when using a corpus of transcribed audio cassette record-

ings, as these usually fail to capture non-verbal response tokens such as head nods and shoulder shrugs⁴. Examples of minimal response tokens include:

- (2) A: Tis a lovely day but tis cold isn't it?
 B: Ah the days are grand shure well yesterday was a bad bad evening.
 A: *Mm*.
 B: It turned black.
 [LCIE]

- (3) A: Her hair is fab isn't it?
 B: Fab?
 A: It's so cool though.
 B: Yeah it's cool all right.
 A: Do you know it's so natural.
 B: *Mm*.
 A: It's a real nice shade like it's not you know.
 [LCIE]

Non-minimal response tokens, on the other hand, include:

- (4) A: I wouldn't have minded giving an apprenticeship to that lad here on the site cos he was a good strong worker so he was. ... he was a polite young fella too.
 B: *Is that right?*
 A: She had a tough job with them she brought up those two kids herself. Her marriage broke down there a long time ago.
 [LCIE]

- (5) A: ... isn't that nice now. Blue sky.
 B: *Lovely*.
 A: A bit of a breeze.
 [LCIE]

As noted by McCarthy (2002), non-minimal response tokens may be pre-modified by intensifying adverbs, which add further emphasis as in the case of *Oh jolly good* below:

- (6) [Woman talking about giving birth]
 A: Dick was very excited cos at one point they asked for hot towels.
 B: Oh.
 A: Just like the movies. So he skipped off down the corridor to get the hot towels.
 B: *Oh jolly good*.
 [CANCODE, McCarthy 2002: 65]

- (7) [Discussing tenancy problems in rented accommodation]
 A: Isn't there something in your tenancy agreement about it? You have a written agreement don't you?
 B: *Most definitely*.
 [CANCODE, McCarthy 2002: 65]

McCarthy (2002) notes that both minimal and non-minimal response tokens can occur in pairs or clusters, as in this example from LCIE:

- (8) A: ... you know it reminds me of am the play and ah.
 B: Mm.
 A: And the character in the play is not+
 B: I don’t know.
 A: +someone I’d kind of identify with+
 B: Yeah that’s true that’s true but I wonder if that’s a cultural sort of+
 A: *Yeah mm*
 B: +I don’t know I had the same question for Rosemary ...
- [LCIE]

Carter & McCarthy (2006) tell us that response token pairings are particularly evident when a topic is being closed down or at a boundary in the talk when another topic is being introduced.

- (9) [Couple asking permission to look at a disused railway line]
 A: It went through, it goes through. Straight, straight on.
 B: *Right. Wonderful. Great.* Can we look round then?
 A: Yes certainly.
 B: Thank you.
- [CANCODE]

McCarthy (2002) and Carter & McCarthy (2006) also point out that the tokens *absolutely*, *certainly* and *definitely* may be negated as response tokens by adding *not*.

- (10) [Speaker A is considering buying a CD player for the first time]
 A: ... but then I’d have to go out and buy lots of CD’s wouldn’t I.
 B: Well yes. I suppose you would.
 A: There’s no point in having a thing if you can’t play them. Haven’t got any.
 B: *Absolutely not. Absolutely not.*
- [CANCODE]

3.2 Functions of listener response tokens

The functions above are used to signal a boundary *and* pragmatically to add satisfaction or agreement, or simply to express friendly social support. Occasionally, triple response tokens occur.

In comparison to the volume of research on forms, relatively few studies address the micro-functions of response tokens in conversation. However, there is enough research available to assert that they have more than one macro discourse function. Yngve (1970), for example, notes from his observations of laboratory conversations, recorded audio-visually, that there is an apparent link between the use of certain forms and the marking of known or common infor-

mation. Mott & Petrie (1995), in line with Bilous & Krauss (1988) and Fishman (1978), point out that listener responses signal support for, or attention to, what the speaker is saying. Fellego (1995), in a study in the context of American English minimal responses, concludes that 94.6% occur at phrase boundaries and that they function both grammatically and socially. Schegloff (1982) identifies the “continuer” function of response tokens. This function will be discussed further below. It refers to how response tokens facilitate the flow of ongoing talk, by signalling listenership at the most basic level. Building on this, Maynard’s (1989) cross-cultural study of Japanese students conversing with American counterparts identified five functions in the data: display of understanding of content; support towards the speaker’s judgement; agreement; strong emotional response; minor addition, correction or request for information. Gardner (1997), who looks at minimal responses in Australian data, points out that each has a distinctive role and interactional function (cf. also Gardner 2002). One of the few studies to look at listener response tokens in a specific social context is Antaki et al. (2000). They use the term *high-grade assessment* to refer to what other studies call *non-minimal responses* in the context of interviews (for example, tokens such as *brilliant*, *excellent*, *smashing* see also Antaki 2002). Antaki et al. (2000) argue that high grade assessments function in a task-oriented rather than in a content-oriented manner within such institutional interactions to mark successful completion of the interactional objective. Though expressed differently, this parallels the findings of McCarthy (2003) that these items function over and above the transactional domain of an interaction.

4. Corpus-based studies of listener response tokens across varieties

4.1 Data and methodology

This study looks at the discourse feature of response tokens in two varieties, in terms of forms and functions. Firstly, in relation to forms, we use two databases of one million words, each extracted from CANCODE (five million words in total, see McCarthy 1998 for further details) and LCIE (just over one million words collected in the Republic of Ireland, but not including Northern Irish English⁵, see Farr et al. 2002), as described above. Each comprises only casual conversation from intimate contexts (that is, friends and families) across the three broad goal types as detailed above. Wordlists and cluster analyses were generated to identify and compare the forms used in the datasets.⁶

Word and cluster lists were generated for both corpora, and from these lists response forms were identified manually by cross-checking qualitatively with transcripts using concordancing. A cut off of the first 5000 items was used. In this selection process, a response token was defined as an item that fills a response slot, but which does not take over the speaker turn. In other words, response tokens are seen as turn yielding. In our analysis, response tokens that

form part of a turn were not included as response tokens. For example, *really* in the following extract was not counted as a response token:

- (11) A: ... basically I think I shouldn’t have gone at all because the prescription he gave me I think I could have gotten over the counter.
 B: *Really?* What did he give you?
 [LCIE]

Whereas *really* in example 12 does count as a response token because it does not take over the speaker turn:

- (12) A: And I don’t think her insurance is even that cheap.
 B: *Really.*
 [LCIE]

We limited our focus on forms to lexicalised items (e.g. *really*, *right*, *absolutely*, *no way*, *oh my God*) in the single word count. Vocalisations (e.g. *mm*, *um-hum*, etc) or other minimal non-lexicalised forms (such as *yep*, *oooh* etc.) were not included as single word tokens.

Finally, in a further step, we extracted from the CANCODE and the LCIE two small, highly comparable corpora in order to examine response token functions. Both of these corpora consisted of 20,000 words of casual conversations, between British and Irish females, all around 20 years of age. All participants were students and close friends, who, in most cases, shared accommodation. These data were examined qualitatively, in terms of all of the response tokens that occurred so as to identify and compare their functions.

4.2 Results

Within the cut-off range of 5000 occurrences of a word or cluster, only items that occurred at least five times as a response token were counted. This yielded 87 tokens in all, 36 in LCIE and 51 in CANCODE.

Table 2. LCIE and CANCODE single-word response tokens within the first 500 words which occurred more than five times (frequency per million words in brackets)⁷

LCIE				CANCODE			
yeah	(2092)*	God	(21)	yeah	(1946)*	lovely	(33)
no	(483)*	lovely	(14)	yes	(1260)	exactly	(23)
right	(268)	maybe	(10)	right	(1200)*	great	(19)
what	(211)*	grand	(9)	no	(989)*	nice	(16)
really	(128)	brilliant	(8)	really	(221)	definitely	(15)
sure	(66)	never	(6)	aye	(241)	never	(11)
Jesus	(57)			okay	(87)	absolutely	(11)
				God	(79)	quite	(8)
				sure	(60)	maybe	(7)
				good	(51)	true	(7)

Table 3. LCIE and CANCODE two-word clusters that occurred more than five times within the 500 most frequent forms (frequency per million words in brackets)

LCIE				CANCODE			
oh yeah	(245)	oh no	(26)	oh yeah	(345)	do you?	(26)
oh right	(220)	all right	(22)	oh right	(244)	oh well	(25)
yeah yeah	(214)	right yeah	(21)	that's right	(166)	do they?	(23)
no no	(45)	Jesus Christ	(13)	oh yes	(122)	isn't it?	(22)
oh God	(40)	my God	(10)	I know	(119)	no no	(17)
is she?	(35)			oh God	(99)	very nice	(16)
				oh dear	(92)	don't you?	(14)
				is it?	(79)	very good	(11)
				all right	(64)	I see	(10)
				did you?	(45)		

Table 4. LCIE and CANCODE three-word clusters that occurred more than five times within the 500 most frequent forms (frequency per million words in brackets)

LCIE				CANCODE			
I don't know	(129)	oh yeah yeah	(20)	I don't know	(93)	do you reckon	(10)
oh my God	(97)	not at all	(12)	oh I see	(44)	I can't remember	(8)
yeah yeah yeah	(63)	oh right yeah	(12)	oh my God	(35)	it doesn't matter	(5)
yeah I know	(46)	no no no	(10)	something like that	(13)		
Are you serious?	(30)	I know that	(8)				
I know yeah	(27)						

Table 5. LCIE and CANCODE four-word clusters that occurred more than five times within the 500 most frequent forms (frequency per million words in brackets)

LCIE				CANCODE			
oh yeah yeah yeah		(21)		I don't think so		(14)	
no no no no		(10)		oh I don't know		(7)	
yeah yeah yeah yeah		(8)		Erm I don't know		(6)	

[No five or six-word clusters occurring with a frequency greater than five were found in either CANCODE or LCIE.]

In terms of comparison at the level of forms, the corpus search brings to light a number of points. Firstly, we see that a broader range of forms is used by British English speakers, at the single and two-word level. Some of the variation in single word forms is attributable to language variety, for example, if we run concordance line searches on the forms that are not common to both LCIE and CANCODE, we find that *grand* (Irish English), and *quite*, yes (British English) are mutually exclusive as response tokens. We also see a broader range of

forms in the British English single-word items which are also found in American English, i.e. *right*, *absolutely*, *sure*, *good*, *lovely*, *exactly*, *great*, *definitely*, *true*, *really* (as noted by McCarthy 2002). In contrast, the Irish single word forms only have *really*, *sure* and *right* in common with McCarthy’s findings for single word non-minimal responses in American English. At a pragmatic level, we note that there are a number of differences. Firstly, *yes* and *quite* in British English have no corresponding occurrence in the Irish data. In the conversation below, two women in their 40s and 50s talk about the speaker A’s chiropractor.

- (13) A: But he’s very nice and what he does is erm <\$=> he doesn’t </\$=> he tries to do the minimum+
 B: Mm.
 A: +to get you right. He doesn’t believe in doing everything all the time.
 B: No.
 A: If things are going well he tries to leave it alone you see.
 [four turns later]
 B: Well the trouble with <unintelligible> who was just so brilliant+
 A: Yeah.
 B: +such such a wonderful man+
 A: *Yes*.
 B: +that everyone seems poor in comparison. [CANCODE]
- (14) A: Well I do hope this Hoover thing is gonna be sorted out. Cos I am not having my flight out to Orlando <\$H> if if poss </\$H> ruined by+
 B: Mm.
 A: +a bunch of Hoover-swinging Scotsmen.
 B: *Quite*. [CANCODE]

We posit that such forms index a higher level of formality in British English. McCarthy (2002) also found occurrences of *quite* in his study of non-minimal forms in British data, and he comments that “intuition and subjective impressions suggest that *quite* as a single word response token is at the very least rather formal in contemporary British speech, and may be on the verge of being perceived as an archaism” (2002: 60). In the CANCODE data in this study, *quite* occurred eight times in one million words of British English.

Other form-related observations which account for the broader spread of two-word items in the British English data include the use of tag questions, such as:

Table 6. Tags in British English, occurrences per million words

Tag	<i>is it?</i>	<i>did you?</i>	<i>do you?</i>	<i>do they?</i>	<i>isn’t it?</i>	<i>don’t you?</i>
-----	---------------	-----------------	----------------	-----------------	------------------	-------------------

Occurrences per million words	79	45	26	23	22	14
-------------------------------	----	----	----	----	----	----

We also found evidence of their use in the Irish data, but only in the form of *is she?*, which occurred 27 times per million words. Carter & McCarthy (2006) use the term *follow-up question* to refer to these forms which they say can function as a signal of engagement and attention by the listener. This function, they note, is often very similar to that of backchannel responses, such as *yeah* and *really*. They support this, saying that follow-up tag questions in informal spoken language often simply function to keep the conversation going by inviting further responses from the listener.

- (15) C: It's one one thing I used to dread.
 A: *Did you?*
 [CANCODE]
- (16) A: [laughing] <background noise> That's polenta. They have polenta all the time on Ready Steady Cook.
 B: *Do they?*
 [CANCODE]

Religious references and swearwords appear in both the British and Irish data. However, their use in the British data is limited to *God* and *oh God*, while the Irish data comprise *God*, *oh God* and *oh my God* and the swear words *Jesus* and *Jesus Christ*. Here is an example from LCIE:

- (17) [Friends are looking at an old school team photo and are trying to identify the people in it]
 A: Ryan the oldest guy Tom Hartnett John Rodgers+
 B: Oh yeah.
 A: +Brian Fitz.
 B: Paul Regan.
 [laughing].
 A: *Jesus Christ*
 B: What year is this?
 A: The late nineties
 [LCIE]

In the differing use and frequency of religious references, we see pragmatic variation that points to a greater level of informality within the Irish data, as the pragmatic impact of *God* is more neutral compared with *Jesus*. However, there is a paradox here which is best understood socio-culturally. The Irish speakers seem to accept swearing as a normal and frequent response token. It seems to have reached semantic neutrality. However, Ireland is still a predominantly Catholic country and so one might expect the opposite to be the case. The explanation may be found in Andersson & Trudgill's (1990) work; they note that

swearing is associated with the areas that are taboo or significant in a particular culture.⁸ Hence, it is because *Jesus* has more significance in Irish society that is it used as a swear word in everyday conversation.

There is also a contrast in the reduplication of forms in both datasets. The Irish data display more reduplication: *yeah yeah, no no, yeah yeah yeah, no no no, oh yeah yeah yeah, no, no, no, no* and *yeah yeah yeah yeah*. For example:

(18) [Three friends are discussing surrogate reproduction]

A: That would kill me seeing someone else having my child.

B: Ah *no no no no no no*. I had this conversation with my mother now.

C: *No no no no no*.

B: No if Caitríona couldn’t have kids or one of my friends or someone and they asked me to have their kid I’d have no problem having it for them.

A: I wouldn’t have a problem doing it but I would have a problem with someone else having it. Imagine having your mother carrying your baby like.

[laughing]

A: My baby would be my sister like.

[laughing]

[LCIE]

The British data, while it has less reduplication, contains more clusters with the vocalisation *oh*:

Table 7. Occurrences of *Oh*-clusters in British English

<i>oh</i> form	<i>oh yeah</i>	<i>oh right</i>	<i>oh yes</i>	<i>oh God</i>	<i>oh dear</i>	<i>oh I see</i>	<i>oh well</i>	<i>oh my God</i>	<i>oh I don't know</i>
Occurrences per million words	345	244	122	99	92	44	25	35	7

For example:

(19) A: They’d been cleaned and put in.

B: Oh they’d put them back in a bag or something had they?

A: No. They weren’t in a bag. They were just inside the chicken.

B: *Oh God*.

A: I just chucked them away and said nowt. [laughing]

[CANCODE]

Finally, we note that the Irish form *Are you serious?* could lead to cross-cultural pragmatic failure (after Thomas 1983) because it could be misunderstood in terms of how the listener orients towards the propositional content of the message. The form, which is used in Irish English as non-minimal response token, is not found in the more dominant variety of British English, and therefore we propose that it has potential for pragmatic confusion or even face threat. Here is an example of its use in Irish English:

- (20) [Three speakers are gossiping about two young men in their locality who have built a house together]
- A: ... he's just built his house it was built in the last six months my god it's a massive yoke the two lads living on their own.
 B: *Are you serious?*
 A: Yes you would be afraid to touch anything.
 B: Aren't they marvellous?
 A: Yeah really like it doesn't look like a home at all cause everything is just perfect.
 C: Like a showhouse.

[LCIE]

5. Functions of response tokens

Hitherto, we have looked at response token forms using relatively large corpus samples (though one million words would be considered a small corpus by contemporary norms). This has allowed us to see lexical patterns using our software. There is no automatic means of extracting and comparing the discourse functions of response tokens, so, in order to overcome this, we have constructed two very small and very comparative datasets which we will examine qualitatively in terms of how response tokens function within them. The datasets are again sub-corpora of LCIE and CANCODE. Both comprise 20,000 words of data and are matched in terms of gender, age, social relationship, two-party to multi-party ratio of interactions, socio-economic class and genre of conversation:

Table 6. Description of sub-corpora of 20,000 words each

Sub-corpus	No. of words	Description
YW20a – Irish English Young women of 20 (LCIE)	20,000	Two sets of 10,000 words (all data are taken from LCIE) 1) of a two-party conversation between close Irish female friends 2) of a multi-party conversation between four close Irish female friends. In all cases the women were students around the age of 20 Topics covered include: gossip about friends and boy-friends, anecdotes and stories.

YW20b – British English Young women of 20 (CANCODE)	20,000	In parallel with sub-corpus YW20a, these data comprise sets of 10,000 words (all data are taken from CANCODE) 1) of a two-party conversation between close British female friends 2) of a multi-party conversation between five close British female friends. In all cases the women were students around the age of 20 Topics covered include: gossip about friends and boy-friends, anecdotes and stories.
---	--------	--

All of the data were read exhaustively so as to manually identify and classify all response tokens. These functional classifications were devised by two raters and cross-checked by a third rater. In terms of frequency of response tokens, we found that there were considerably more in the British data. This is in line with the finding above that fewer response token types were used in Irish English. This result allows us to speculate that there is more response token use in British English than in Irish English. However, this hypothesis would merit a separate investigation.

Table 7. Frequency of forms in YW20a (Irish) and YW20b (British) datasets

Corpus	YW20a (Irish)	YW20b (British)
Frequency/20,000 words	191	304

Before we go into greater quantitative detail in our comparison, we first outline the four broad functions which we identified in these data as a whole:

Table 8. Functions of response tokens in casual conversation

Type of token	Function	Typical examples
Continuer tokens*	Maintain the flow of the discourse.	Minimal forms such as <i>Yeah, mm.</i>
Convergence tokens	Markers of agreement/convergence. They are linked to points in the discourse: 1) where there is a topic boundary or closure 2) where there is a need to converge on an understanding of what is common ground or shared knowledge between participants.	Many forms can perform this function such as: • single word items: <i>yeah</i> • follow-up questions such as <i>did you?, is she?</i> • short statements, e.g. agreeing statements: <i>yeah it's pretty sad.</i>

Engagement tokens	Markers of high engagement where addressee(s) respond on an affective level to the content of the message. These backchannels express genuine emotional responses such as surprise, shock, horror, sympathy, empathy and so on.	They manifest in many forms for example: <ul style="list-style-type: none"> • single-word forms, such as <i>excellent, absolutely</i> • short statements, repetitions: <i>that's nice, oh wow, oh really</i> • follow-up questions: <i>did you?</i>
Information receipt tokens	Markers of points in the discourse where adequate information has been received. These responses can impose a boundary in the discourse and can signal a point of topic transition or closure, and they can be indicative of asymmetrical discourse.	<i>Right and okay</i>

* (after Schegloff 1982)

5.1 Continuer response tokens

Continuer response tokens are facilitative in that they maintain the flow of talk. As the term suggests, they encourage the current speaker to continue. As mentioned above, many researchers have identified this function of listener response and noted that it is usually realised using minimal response tokens (see Schegloff 1982, Maynard 1989, Gardner 1997, 1998, 2002). Speakers perceive continuer response tokens as floor-yielding signals that mark the addressee's desire for the talk to continue. An analysis of concordance lines for a minimal response token such as *mm*, reveals that it is surrounded by ongoing utterances rather than being part of a turn itself. In the extract (21) taken from the LCIE YW20a corpus, a friend is telling of a text message "conversation" she had with her boyfriend (note *messing* is Irish English slang for *joking*). *Yeah* signals that the listener is eager for the story to continue:

- (21) A: And he sent one back saying "ah come on now Sinead are you messing or are you serious like?"
 B: *Yeah*
 A: And ah he sent one saying "no I'm deadly serious am I'm going to kill you when I catch you" so the next thing your man was pure upset over this like and...

[LCIE – YW20a]

We can observe from this example that continuer tokens are facilitative in that they maintain the flow of talk. They may be perceived by the speaker as floor-yielding signals that mark the addressee's desire for the narrative to continue. In the following example from CANCODE YW20b in which friends are talking about buying a pair of shoes, the minimal response token *mm* facilitates the flow of the conversation.

- (22) [*Superdrug* refers to the name of a British shop]
- A: I didn’t even know they sold erm shoes in there.
 B: No. I didn’t know they sold shoes.
 C: Didn’t know that.
 A: But erm.
 C: They’re really nice.
 A: Cos like it’s really weird cos I had erm you know when you think of something you want to have.
 C: *Mm.*
 A: And you haven’t seen them in the shops.
 C: *Mm.*
 A: I sort of thought oh I really want you know. And I sort of visualised what I wanted and then erm I went down *Superdrug* with Rachel and we popped in and I thought Ooh. They’re the ones I want.
 [CANCODE – YW20b]

5.2 Convergence response tokens

Close examination of the corpus examples shows that response tokens are most frequently found at points of convergence in conversations, that is, where participants agree, or simply converge on opinions or mundane topics and this leads them to negotiate topic boundary points collaboratively, where a topic can be shifted or changed. Convergence can also be followed by a conversational closure point. In this way, response tokens have a pragmatic function in that they help bring about agreement and convergence leading sometimes to topic shifts. In the following example from the CANCODE excerpt YW20b between female flatmates, we see that the topic (a great night out that the friends had together) has run its course and it is collaboratively rounded off with the non-minimal response token *you never know*. Notice also how this phrase is a recycling of a phrase from the previous turn, which makes for a very symmetrical ending point at which participants converge topically and lexically before moving on to a new topic:

- (23) A: Yeah. We haven’t had a night like that for a while have we?
 B: No. Must have another one.
 A: Silly night. [laughing] What?
 B: Must have another one.
 A: Well I think we will.
 B: Wednesday.
 A: *Mm.* Lifts the spirits.
 B: *Mm.* *You never know* we might be able to get a new recruit.
 A: [laughing] *You never know.*
 B: [laughing]
 [CANCODE – YW20b]

After this point in the conversation, the topic shifts. In example 24 below, two close friends are chatting about a former classmate who committed suicide. We

see how one phase of the narrative ends with an evaluative formulation phased over two turns: *it just goes to show you can't take people at face value* and *And you don't know what's going on*. This evaluation is unchallenged by the addressee and convergence is signalled after each phase of the evaluation by the response tokens *no* and *exactly*. This registers the addressee's agreement and it allows the conversation to move to a side sequence to this tragic story (see turns 5 and 6) to which both participants contribute:

- (24) A: ... it just goes to show you can't take people at face value.
 B: *No*.
 A: And you don't know what's going on either.
 B: *Exactly*.
 A: But am seemingly she knew what she was doing as well because she brought the+
 B: Oh she had it all planned out. She brought the little brother into get a present inside in Galway...

[LCIE – YW20a]

Adolphs & O'Keeffe (2002) note that as well as helping to bring about topic shifts, these tokens are often found in closings as they allow conversations to come to a collaborative end. The authors illustrate how in an Irish radio phone-in show, *Liveline*, the presenter uses them and other markers of agreement in the closing of the call:

- (25) [The presenter and the caller are chatting about the merits of clip on earrings]
 Presenter: And aren't they grand?
 Caller: Yes they're very very handy.
 Presenter: *Yeah*.
 Caller: But they're not as secure as having them in your ear.
 Presenter: *This is true. This is true*.
 Caller: You know you could lose them easily.
 Presenter: *That's true*. O.K. Tess well thanks for talking to us thanks very much
 Caller: Right thanks very much. Bye
 Presenter: All the best. Thank you indeed bye bye bye bye.

[LCIE]

McCarthy (2003) notes that non-minimal responses tokens sometimes cluster in consecutive series across speakers, providing multiple signals that a conversation is about to be terminated, while at the same time consolidating interpersonal relationships. He also observes that they often occur together with other markers of closure, such as thanks, checks, confirmations and greetings, and that clustering is especially frequent in telephone conversations where there are often pre-closing and closing routines.

In a pragmatic sense, the affective value of convergence response tokens is worth noting. These tokens are of higher relational value than continuer tokens

(see above). They do more than just signal turn-yielding, listenership and a desire for the narrative to continue. Signalling agreement or converging on mundane topics is a form of interactional bonding between speaker and addressee and convergence response tokens help maintain good relations between speakers by reinforcing commonality between them.

5.3 Engagement tokens

This type of response token again functions very much at an affective level. Engagement tokens thus signal the addressee’s enthusiasm, empathy, sympathy, surprise, shock and disgust at what the speaker is saying, without taking over the turn. They are also indicative of the addressee’s high level of engagement with the content of the speaker’s message. These tokens are typically non-minimal responses and common items include *brilliant*, *absolutely*, *wow*, *cool*, *gosh*, *really* and short phrases, such as *that’s tough*, *you’re not serious*, *Is that so?* In example (26), an engagement token is used to express the addressee’s delight at what her friend is saying. Speaker B is talking about how she will spend the summer with her boyfriend in Edinburgh (note: Debenhams is a well-known British department store; CV refers to curriculum vitae or résumé):

- (26) A: What are you going to do about a job?
 B: I don’t know. He says that it’s going to be like Killarney and that I should get one easily enough and I’ve been in contact with Debenhams and they told me to send over my CV.
 A: *Brilliant Mary brilliant.*
- [LCIE – YW20a]

In Example (27), we see an engagement token signalling the addressee’s sympathy with the speaker’s message using a vocalisation.

- (27) [Speaker A has just told a story of how she and her boyfriend had a row a few days earlier]
- A: Were you out last night?
 B: I was.
 A: Where were you?
 B: Am you see we had to reconcile last night and get it all back on.
 A: *Aaahhhh*
 B: He says to me “I forgive you anyway” he says “for what you did to me” and I says “I was only testing the waters”.
- [LCIE – YW20a]

Examples from the British YW20b data include:

- (28) A: He was singing to me. [laughing] And then he come over and he gave me one. And he gave me a peck on the he kissed me on the forehead and gave me a hug.

- B: *Ah. That's nice.*
 A: [laughing] And then walked me home. It was hand in hand skipping up the road. [laughing] And gave me a hug goodbye.
 B: I've had flowers given to me.
 A: *That was nice.*
- [CANCODE – YW20b]

- (29) A: I ate almost a whole jar of Roses this weekend.
 B: *Did you?*
 [laughing]
- [CANCODE – YW20b]

This type of response token functions at a much higher relational level than continuer tokens since they not only signal a desire for the speaker to continue, but also communicate the addressee's affective response to the speaker's message.

5.4 Information receipt tokens

We found a small number of response tokens in both datasets which did not fit any of the above categories. While the previous three types of tokens seemed to serve relational functions, a further type of token seemed to have a more organisational function. These backchannels were usually also marked by falling pitch. In the few examples that we found, they seemed to serve a global discourse marking function (cf. Lenk 1998) within the orientation stage of a narrative. This response token is used as a "self-imposed" pragmatic marker at which the storyteller marks a boundary where the narrative can begin now that the contingent details are clear for the participants. In example (30), taken from the Irish casual conversation data (YW20a), we see that when the storyteller uses an information receipt token at the point where she assumes all of the contingent details are in place to continue with the story, the listener signals that she is not ready and still needs more details (or at least confirmation of an assumed piece information). The storyteller provides these before continuing.

- (30) A: He's been in Wexford for years right. I told you he's separated didn't I? And that he has a child.
 B: Yeah.
 A: *Right.*
 B: But he's only young isn't he?
 A: He's only 29.
- [LCIE – YW20a]

In extract (31) from the British data (YW20b), we see another instance of an information receipt token being "self-imposed" so as to organize information in the pre-amble to an anecdote. Here we see that speaker C signals that she is au

fait with the contingent details, but this is not the case for B and so there is a prolonged stage of inquiry about the character of the forthcoming anecdote:

- (31) A: ... I just saw this person I thought was quite nice but I can’t remember what he looks like ... [laughing]
 B: Is that Trinity?
 A: Mm.
 C: *Oh right*.
 B: Is he tall? Short?
 A: Don’t know really.
 ?: <unintelligible>
 B: Distorted.
 [laughing]
 A: No. He’s not like really short and he’s not really tall. He’s sort of.
 B: Average height.
 A: Average. Yeah.
 [laughing]
 B: A normal sort of guy.
 [laughing]

[CANCODE – YW20b]

Adolphs & O’Keeffe (2002) and O’Keeffe (2003) found this type of response token, particularly in the form of *right*, to be very prevalent in their analyses of Irish radio phone-in data. Here, the presenter used such response tokens in an organisational manner. Adolphs & O’Keeffe propose that this token is strongly associated with asymmetrical interaction where one of the participants is a power role holder (see below for a functional comparison, Figure 3). McCarthy (2003) has also noted that some response tokens are strongly associated with particular contexts. *Fine*, he suggests, most typically occurs in making arrangements and reaching decisions and *certainly* most typically occurs in reply to a request for a service or favour:

- (32) A: Okay. I’ll see you a bit later then.
 B: *Fine*.
 A: In the morning, whenever.

[CANCODE]

- (33) [To a waiter]
 A: Can I have the bill please?
 B: *Yes, certainly*.

[CANCODE]

McCarthy (2003) notes that adjectives such as *excellent*, *fine*, *great*, *good*, *lovely*, *right*, *perfect* offer positive feedback to the speaker and often mark the boundaries of topics, where speakers express their satisfaction with phases of business such as making arrangements, agreeing on courses of action, and marking the satisfactory exchange of information, goods and services.

(34) [At a travel agent's]

Assistant: There you go. There's your ticket. And your accommodation there. Insurance, and just some general information.

Customer: *Excellent. Right.*

[CANCODE]

(35) [Dealer (A) and customer (B) in a car spare parts depot]

A: I'll get one of the lads in to come and do it for you.

B: *Lovely.*

[CANCODE]

When we look at the functions quantitatively, we find that the pattern of their distribution is reasonably similar in the Irish and British young women data, with the function of *convergence* tokens being the most frequent followed by *engagement* tokens.

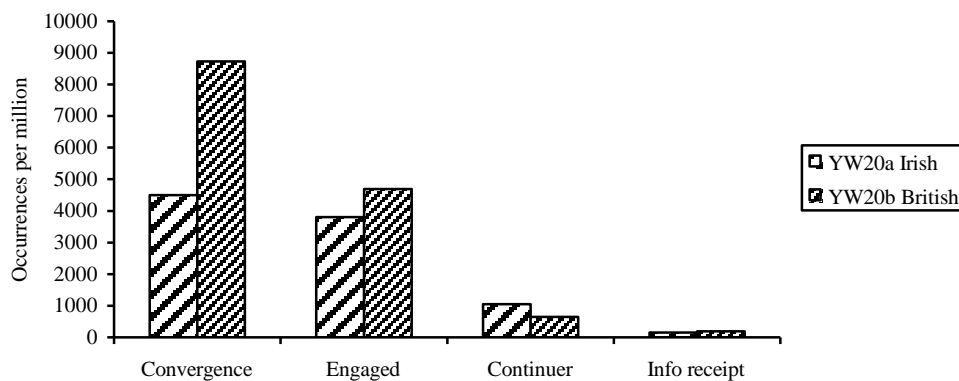
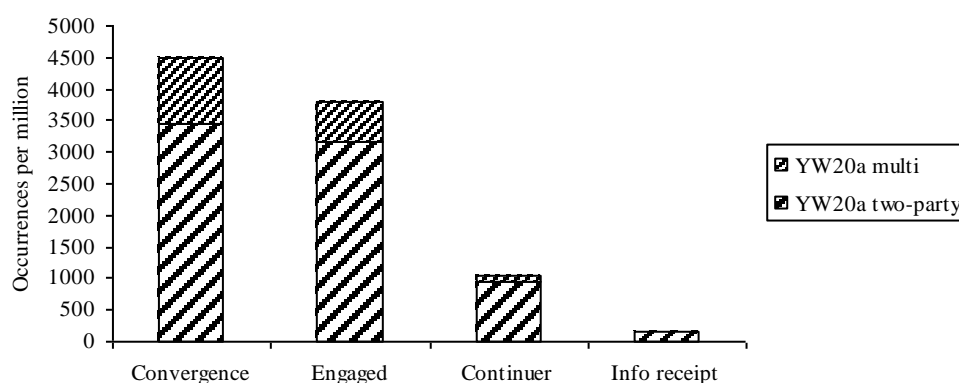


Figure 1. Functional distribution of response tokens in British and Irish young woman data (two-party and multi-party). Results in occurrences per million words.

However, when the results were compared across multi- versus two-party interactions (below), we found them to be inversely proportional. In the Irish data, there are far more response tokens used in the two-party conversation (see wide-striped bar portion of Figure 2a) while in the British data, this result is reversed. It is not possible to draw any broad conclusions from this finding without conducting further analyses. The functional pattern of distribution re-



mained the same, with convergence tokens remaining the most frequent type of response token used in all categories:

Figure 2a. Profile of functions in YW20a (Irish) two-party and multi-party conversations (results in occurrences per million words)

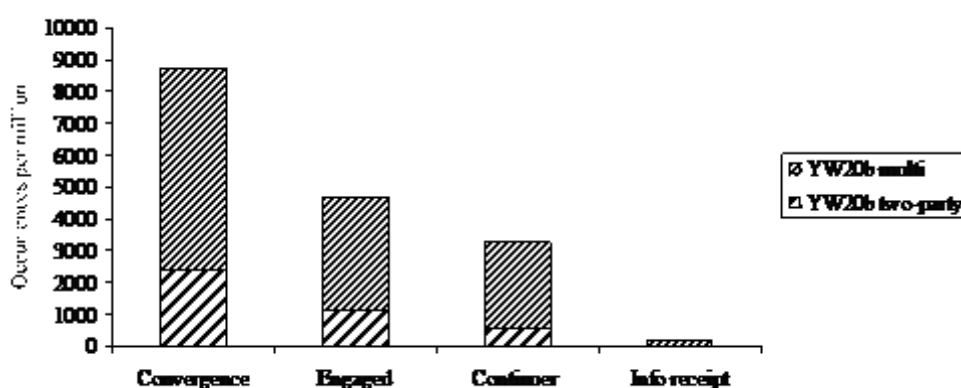


Figure 2b. Profile of functions in YW20b (British) two-party and multi-party conversations (results in occurrences per million words)

The YW20a and YW20b datasets were closely matched in terms of age, gender, social relationship, socio-economic class, but we posit that the type of conversation (everyday conversations about friends, shopping, boyfriends etc.) is the most influential factor resulting in the homogeneity of functional distribution. This is perhaps substantiated by a comparison with our earlier findings (Adolphs & O’Keeffe 2002, O’Keeffe 2003), when we conducted a similar functional analysis of 20,000 words of interactions from an Irish radio phone-in show, *Liveline*. Figure 3 compares the functions of response tokens in the young women’s data and the Irish radio phone-in corpus.

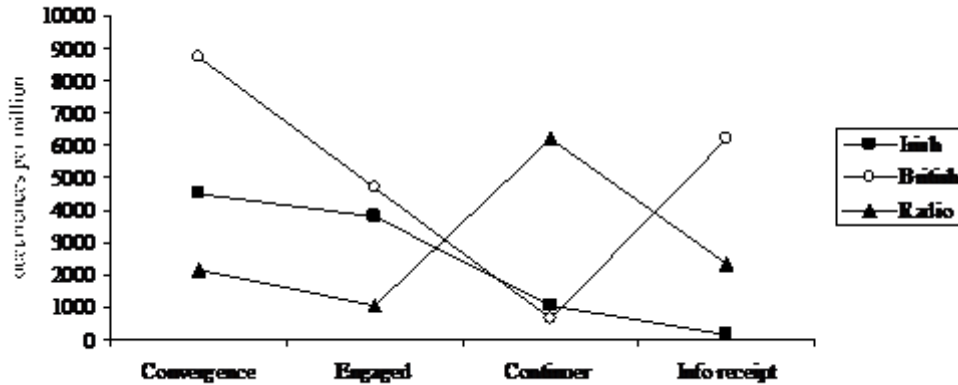


Figure 3. Comparative of functions in YW20a and b with Irish radio phone-in functions across 20,000 words (results presented per million words)

Because the genre of conversation differs, we find a different functional pattern. Most noticeably, the substantially higher frequency of response tokens which function as continuers in radio discourse. This is attributable also to the mode of communication as radio conversations take place in sound-only mode.

6. Conclusion

We set out to address the dearth of comparative pragmatic research in terms of how varieties of English differ. Within the paradigm of variational pragmatics, we also wanted to test corpus linguistics as a methodological tool. The focus of our study was the discourse feature of listener response. Firstly, we can say that the area of variational pragmatics, from the perspective of our study, has great potential for development, even within this one area of listener response tokens. Further studies are planned that will look at how listener response tokens differ pragmatically in American English, as well as in languages other than English. The paradigm of variational pragmatics, therefore, serves us well and will form the basis of much more of our future comparative work.

From this short study alone, it is clear that even though British and Irish English are two neighbouring varieties of English with frequent contact, they do not represent a single monolithic entity. Rather, variation is found at the level of form, these forms reflecting socio-cultural norms and subtleties that differ in Irish and British society. We also found considerable differences in the frequency of listener response token activity. British English conversations contained far more. This is something that merits further investigation and raises questions such as: Are British people better listeners? Do Irish people talk more and respond less? Do Irish people yield turns less and interrupt more? We also noted the use and variation in forms which involved religious refer-

ence, or which were swear words. These prevailed more in the Irish data. At a socio-cultural level, we speculated that their higher frequency in Irish English was linked to the role of religion in Irish society. Hence, they have greater taboo value in that variety.

In our quest to compare the data functionally, we undertook a qualitative study. The results from this manual analysis of two sub-corpora of 20,000 words (each of closely matched data with respect to gender, age, social relationship, socio-economic class and genre of discourse) pointed to three main functions of response tokens in this context, and also to a minor function, *information receipt* marking. At the level of overall frequency, we again found a discrepancy between British and Irish English, where the British data contained 59% more response tokens. However, at the level of response token function, we did not find there to be any difference between their use in British and Irish English in these cohorts. In addition, the overall functional pattern held constant for two-party and multi-party interactions in both datasets. This leads us to assert, therefore, that while we have observed differences in the forms, frequencies and sociocultural subtleties of response tokens in British and Irish English, the pragmatics of the discourse function itself appear to be constant. We note also that this manual phase of the study only looked at female data (in order to control the feature of gender). In the future, we hope to replicate this study using male data, where all other variables are controlled.

General statements about response tokens that arise from this study include the following:

- Response tokens are core fluency items which function pragmatically to show listenership
- The items are discourse tokens rather than adverbs or adjectives
- A vocabulary of non-minimal response tokens probably exists in all languages, but seems to vary within and between languages
- Even between language varieties there is potential for cross-cultural pragmatic failure

Finally, let us consider corpus linguistics as a methodological tool for the study of variational pragmatics. It has proved very useful to us as a means of accessing large amounts of spoken language samples, which we have easily been able to control for a number of variables. It allowed us to automatically retrieve results and also compare forms in the datasets (in total amounting to two million words, approximately 170 hours of talk). Where there was need to disambiguate forms or identify only those forms which functioned as response tokens, we had computerised access to the source files and the exact location in the original conversations in which the items occurred (as well as to all of the speaker information for that conversation). In this sense, it is undoubtedly a tool of considerable merit. In terms of its limitations, firstly, it is only a tool and requires other frameworks for the interpretation of data. For example, discourse analysis, conversation analysis and pragmatics aided our understanding of the

forms and functions of response tokens, and our work is based on a long lineage of research in these areas. Secondly, spoken corpora are only beginning to be digitised. We are working with transcriptions of audio recordings. They have gone from the moment of recording to the person who transcribed them before our research. This raises a number of issues: 1) they have been extracted from their audio-visual situational context and transposed into the written word, 2) in so doing they have lost much of their prosodic integrity, as well as 3) visual clues such as head nods (which could operate as surrogate response tokens) and facial expressions and 4) though we can do so much automatically, corpus data still require manual and qualitative work to offset and interpret purely quantitative data. The future of spoken corpora is with digital audio-visual recording, where sound and image can be aligned to transcriptions, and we are now at a stage where technology can allow for this. With this in mind, the potential of corpus linguistics as a tool to aid our understanding of variational pragmatics is very promising.

Notes

1. The authors wish to thank Dawn Knight who worked as research assistant on this paper.
2. CANCODE was built by Cambridge University Press and the University of Nottingham and it forms part of the Cambridge International Corpus (CIC). It provides insights into language use, and offers a resource to supplement what is already known about English from other, non-corpus-based research, thereby providing valuable and accurate information for researchers and those preparing teaching materials. Sole copyright of the corpus resides with Cambridge University Press, from whom all permission to reproduce material must be obtained.
3. Most data is copyright so ensure that clearance or permission is sought beforehand.
4. Recent multi-media corpus projects may, however, be able to obviate this problem by the use of synchronised video records alongside the conversational transcript. See, for example, Reder et al. (2003).
5. Donegal, for example, while part of the Republic of Ireland was not included as Northern Irish English is used there.
6. Wordlist generation is a core corpus software function which facilitates the rank ordering of all the words in order of frequency. Cluster analysis is similar to this process, except that it looks for clusters of words as opposed to single word items, for example, two-word clusters (*you know*), three-word clusters (*Are you sure?*), four-word clusters (*know what I mean?*) and so on.
7. An asterisk marks results based on a random selection of 5000 occurrences. The figures without an asterisk represent items which occurred less than 5000 times in the corpus. In the latter case, all items which occurred in the corpus were analysed. In the former case, only a sample 5000 occurrences could be analysed due to very high overall occurrences (e.g. 10,000 occurrences) which would have made the analysis unmanageable.
8. We are grateful to Brona Murphy for pointing this out.

References

- Adolphs, Svenja (2006). *Introducing Electronic Text Analysis: A Practical Guide for Language and Literary Studies*. London: Routledge.
- Adolphs, Svenja, Brown, Brian, Carter, Ronald, Crawford, Paul & Sahota, Opinder (2004). “Applying corpus linguistics in a health care context”. *Journal of Applied Linguistics*, 1(1), 9–28.
- Adolphs, Svenja & O’Keeffe, Anne (2002). “Response in British and Irish English: ‘Go away!’” Paper presented at the First IVACS International Conference, University of Limerick, 14–15 June, 2002.
- Aijmer, Karin & Altenberg, Bengt (Eds.). (1991). *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. London: Longman.
- Andersson, Lars & Trudgill, Peter (1990). *Bad Language*. Oxford: Basil Blackwell.
- Antaki, Charles (2002). “‘Lovely’: Turn-initial high-grade assessments in telephone closings”. *Discourse Studies*, 4, 5–24.
- Antaki, Charles, Houtkoop-Steenstra, Hanneke & Rapley, Mark (2000). “‘Brilliant. Next question...’: High-grade assessment sequences in the completion of interactional units”. *Research on Language and Social Interaction*, 33, 235–262.
- Aston, Guy & Burnard, Lou (1998). *The BNC Handbook: Exploring the British National Corpus with SARA*. Edinburgh: Edinburgh University Press.
- Atkins, Sue, Clear, Jeremy & Ostler, Nicholas (1992). “Corpus design criteria”. *Literary and Linguistic Computing*, 7(1), 1–16.
- Banjo, Ayo (1996). “The sociolinguistics of English in Nigeria and the ICE project”. In S. Greenbaum (Ed.), *Comparing English Word-Wide: The International Corpus of English* (pp. 239–248). Oxford: Oxford University Press.
- Barlow, Michael (2000). *Corpus of Spoken, Professional American English*. Houston, TX: Athelstan.
- Baynham, Michael (1991). “Speech reporting as discourse strategy: Some issues of acquisition and use”. *Australian Review of Applied Linguistics*, 14, 87–114.
- Baynham, Michael (1996). Direct speech: What’s it doing in non-narrative discourse? *Journal of Pragmatics*, 25, 61–81.
- Biber, Douglas (1993). “Representativeness in corpus design”. *Literary and Linguistic Computing*, 8(4), 243–257.
- Biber, Douglas, Conrad, Susan & Reppen, Randi (1998). *Corpus Linguistics Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Biber, Douglas, Conrad, Susan, Finegan, Edward, Johansson, Stig & Leech, Geoffrey (1999). *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Bilous, Frances R. & Krauss, Robert M. (1988). “Dominance and accommodation in the conversational behaviours of same- and mixed-gender dyads”. *Language and Communication*, 8, 183–194.
- Bolton, Kingsley, Gisborne, Nikolas, Hung, Joseph & Nelson, Gerald (2003). *The International Corpus of English Project in Hong Kong*. Amsterdam: John Benjamins.
- Boxer, Diana, & Pickering, Lucy (1995). “Problems in the presentation of speech acts in ELT materials: The case of complaints”. *ELT Journal*, 49, 99–158.
- Carter, Ronald, A. (1998). “Orders of reality: CANCODE, communication and culture”. *ELT Journal*, 52, 43–56.
- Carter, Ronald, A. & McCarthy, Michael, J. (1995). “Grammar and the spoken language”. *Applied Linguistics*, 16(2), 141–158.
- Carter, Ronald, A. & McCarthy, Michael, J. (2006). *The Cambridge Grammar of English*. Cambridge: Cambridge University Press.
- Charteris-Black, Jonathan (2004). “Why ‘an angel rides in the whirlwind and directs the storm’: A corpus-based comparative study of metaphor in British and American political discourse”. *Language and Computers*, 49(1), 133–150.

- Cheng, Winnie & Warren, Martin (1999). "Facilitating a description of intercultural conversations: The Hong Kong Corpus of Conversational English". *ICAME Journal*, 23, 5–20.
- Cheng, Winnie & Warren, Martin (2000). "The Hong Kong Corpus of Spoken English: Language learning through language description". In L. Burnard & T. McEnery (Eds.), *Rethinking Language Pedagogy from a Corpus Perspective* (pp. 133–144). Frankfurt a.M.: Peter Lang.
- Cheng, Winnie & Warren, Martin (2002). "'// ㄣ beef ball // → you like //: The intonation of declarative-mood questions in a corpus of Hong Kong English". *Teanga*, 21, 151–165.
- Coates, Jennifer (1986). *Women, Men and Language: A Sociolinguistic Account of Sex Differences in Language*. London: Longman.
- Cotterill, Janet (2004). "Collocation, connotation, and courtroom semantics: Lawyers' control of witness testimony through lexical negotiation". *Applied Linguistics*, 25(4), 513–537.
- Crowdy, Steve (1994). "Spoken corpus transcription". *Literary and Linguistic Computing*, 9, 25–28.
- De Cock, Sylvie (1998). "A recurrent word combination approach to the study of formulae in the speech of native and non-native speakers of English". *International Journal of Corpus Linguistics*, 3, 59–80.
- De Cock, Sylvie (2000). "Repetitive phrasal chunkiness and advanced EFL speech and writing". In C. Mair & M. Hundt (Eds.), *Corpus Linguistics and Linguistic Theory: Papers from the Twentieth International Conference on English Language Research on Computerized Corpora (ICAME 20), Freiburg im Breisgau 1999* (pp. 51–68). Amsterdam: Rodopi.
- Drummond, Kent & Hopper, Robert (1993a). "Some uses of 'yeah'". *Research on Language and Social Interaction*, 26, 203–312.
- Drummond, Kent & Hopper, Robert (1993b). "Backchannels revisited: Acknowledgement tokens and speakership incipency". *Research on Language and Social Interaction*, 26, 157–177.
- Du Bois, John W., Chafe, Wallace L., Meyer, Charles & Thompson, Sandra A. 2000. *Santa Barbara Corpus of Spoken American English*. Philadelphia: Linguistic Data Consortium.
- Duncan, Starkey D. Jr. & Niederehe, George (1974). "On signaling that it's your turn to speak". *Journal of Experimental Social Psychology*, 10, 234–247.
- Farr, Fiona (2002). "Classroom interrogations: How productive?" *The Teacher Trainer*, 16(1), 19–23.
- Farr, Fiona (2003). "Engaged listenership in spoken academic discourse: The case of student-tutor meetings". *Journal of English for Academic Purposes*, 2(1), 67–85.
- Farr, Fiona, Murphy, Bróna & O'Keeffe, Anne (2002). "The Limerick Corpus of Irish English: Design, description and application". *Teanga*, 21, 5–29.
- Fellego, Anna M. (1995). "Patterns and functions of minimal response". *American Speech*, 70, 186–199.
- Fishman, Pamela M. (1978). "Interaction: The work women do". *Social Problems*, 25, 397–406.
- Fries, Charles C. (1952). *The Structure of English*. New York: Harcourt, Brace & Co.
- Gardner, Rod (1997). "The listener and minimal responses in conversational interaction". *Prospect*, 12, 12–32.
- Gardner, Rod (1998). "Between speaking and listening: The vocalization of understandings". *Applied Linguistics*, 19, 204–224.
- Gardner, Rod (2002). *When Listeners Talk: Response Tokens and Listener Stance*. Amsterdam: John Benjamins.
- Granger, Sylviane, Hung, Joseph & Petch-Tyson, S. Stephanie (Eds.). (2002). *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins.
- Holmes, Janet (1988). "Doubt and certainty in ESL textbooks". *Applied Linguistics*, 9, 21–44.

- Holmes, Janet (1996). "The New Zealand spoken component of ICE: Some methodological challenges". In S. Greenbaum (Ed.), *Comparing English World-Wide: The International Corpus of English* (pp. 163–178). Oxford: Oxford University Press.
- Hughes, Rebecca & McCarthy, Michael J. (1998). "From sentence to discourse: Discourse grammar and English language teaching". *TESOL Quarterly*, 32, 263–287.
- Ide, Nancy & Macleod, Catherine (2001). "The American National Corpus: A standardized resource of American English". In P. Rayson, A. Wilson, T. McEnery, A. Hardie & S. Khoja (Eds.), *Proceedings of Corpus Linguistics 2000* (Vol. 13, pp. 274–280). Lancaster: University of Lancaster.
- Ide, Nancy, Reppen, Randi & Suderman, Keith (2002). "The American National Corpus: More than the web can provide". In *Proceedings of the Third Language Resources and Evaluation Conference (LREC), Las Palmas, Spain, 28–31 May, 2002* (pp. 839–844). Available from <<http://americannationalcorpus.org/pubs.html>>.
- Kallen, Jeffery L. & Kirk, John M. (2001). "Convergence and divergence in the verb phrase in Irish Standard English: A corpus-based approach". In J. M. Kirk & D. P. Ó Baoill (Eds.), *Language Links: the Languages of Scotland and Ireland* (pp. 59–79). Belfast: Cló Ollscoil na Banríona.
- Kendon, Adam (1967). "Some functions of gaze-direction in social interaction". *Acta Psychologica*, 20, 22–63.
- Kennedy, Graeme (1998). *An Introduction to Corpus Linguistics*. London: Longman.
- Kettemann, Bernhard (1995). "Concordancing in English language teaching". *TELL and CALL*, 4, 4–15.
- Koester, Almut (2006). *Investigating Workplace Discourse*. London: Routledge.
- Lenk, Uta (1998). "Discourse markers and global coherence in conversation". *Journal of Pragmatics*, 30, 245–257.
- Maynard, Senko K. (1989). *Japanese Conversation: Self-Contextualization through Structure and Interactional Management*. Norwood, NJ: Ablex.
- Maynard, Senko K. (1990). "Conversation management in contrast: Listener response in Japanese and American English". *Journal of Pragmatics*, 14, 397–412.
- Maynard, Senko K. (1997). "Analysing interactional management in native/non-native English conversation: A case of listener response". *IRAL*, 35, 37–60.
- McCarthy, Michael J. (1998). *Spoken Language and Applied Linguistics*. Cambridge: Cambridge University Press.
- McCarthy, Michael J. (2002). "Good listenership made plain: Non-minimal response tokens in British and American spoken English". In R. Reppen, S. Fitzmaurice & D. Biber (Eds.), *Using Corpora to Explore Linguistic Variation* (pp. 49–72). Amsterdam: John Benjamins.
- McCarthy, Michael J. (2003). "Talking back: 'Small' interactional response tokens in everyday conversation". *Research on Language and Social Interaction*, 36, 33–63.
- McCarthy, Michael J., & Carter, Ronald A. (2000). "Feeding back: Non-minimal response tokens in everyday English conversation". In C. Heffer & H. Sauntson (Eds.), *Words in Context: A Tribute to John Sinclair on His Retirement* (pp. 263–283). Birmingham: University of Birmingham.
- McCarthy, Michael J. & Carter, Ronald A. (2002). "From conversation to corpus: A dual analysis of a broadcast political interview". In A. Sánchez-Macarro (Ed.), *Windows on the World: Media Discourse in English* (pp. 15–39). Valencia: University of Valencia Press.
- McCarthy, Michael J. & O'Keeffe, Anne (2004). "Research in the teaching of speaking". *Annual Review of Applied Linguistics*, 24, 26–43.
- Mott, Helen & Petrie, Helen (1995). "Workplace interactions: Women's linguistic behaviour". *Journal of Social Psychology*, 14, 324–336.
- Nelson, Gerald, Wallis, Sean & Aarts, Bas (2002). *Exploring Natural Language: Working with the British Component of the International Corpus of English*. Amsterdam: John Benjamins.
- Nero, Shondel (2000). "The changing faces of English: A Caribbean perspective". *TESOL Quarterly*, 34(3), 483–510.

- O'Keeffe, Anne (2002). "Exploring indices of national identity in a corpus of radio phone-in data from Irish radio". In A. Sánchez-Macarro (Ed.), *Windows on the World: Media Discourse in English* (pp. 91–113). Valencia: University of Valencia Press.
- O'Keeffe, Anne (2003). "Strangers on the line: A corpus-based lexicogrammatical analysis of radio phone-in". Ph.D. dissertation, University of Limerick.
- O'Keeffe, Anne (2006). *Investigating Media Discourse*. London: Routledge.
- O'Keeffe, Anne & Breen, Michael (in press). "At the hands of the brothers: a corpus-based lexicogrammatical analysis of stance in newspaper reporting of child sexual abuse cases". In J. Cotterill (Ed.), *The Language of Sexual Crime*. Basingstoke: Palgrave.
- O'Keeffe, Anne & Farr, Fiona (2003). "Using language corpora in language teacher education: Pedagogic, linguistic and cultural insights". *TESOL Quarterly*, 37(3), 389–418.
- Ooi, Vincent (1997). "Analysing the Singapore ICE Corpus for lexicographic evidence". In M. Ljung (Ed.), *Corpus-Based Studies in English* (pp. 245–260). Amsterdam: Rodopi.
- Reder, Stephen, Harris, Kathryn & Setzler, Kristen (2003). "The multimedia adult ESL learner corpus". *TESOL Quarterly*, 37(3), 546–557.
- Renouf, Antoinette (1997). "Teaching corpus linguistics to teachers of English". In A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (Eds.), *Teaching and Language Corpora* (pp. 255–266). New York: Longman.
- Roger, Derek, Bull, Peter & Smyth, Sally (1988). "The development of a comprehensive system for classifying interruptions". *Journal of Language and Social Psychology*, 7, 27–34.
- Schegloff, Emanuel A. (1982). "Discourse as interactional achievement: Some uses of 'uh huh' and other things that come between sentences". In D. Tannen (Ed.), *Analysing Discourse: Text and Talk* (pp. 71–93). Washington, D.C.: Georgetown University Press.
- Seidlhofer, Barbara (2001). "Closing a conceptual gap: The case for a description of English as a lingua franca". *International Journal of Applied Linguistics*, 11, 133–158.
- Simpson, Rita, Lucka, Bret & Ovens, Janine (2000). "Methodological challenges of planning a spoken corpus with pedagogical outcomes". In L. Burnard & T. McEnery (Eds.), *Rethinking Language Pedagogy from a Corpus Perspective: Papers from the Third International Conference on Teaching and Language Corpora* (pp. 43–49). Frankfurt: Peter Lang.
- Sinclair, John (1997). "Corpus evidence in language description". In A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (Eds.), *Teaching and Language Corpora* (pp. 27–39). London: Longman.
- Stern, Karen (1997). "The Longman Spoken American Corpus: Providing an in-depth analysis of everyday English". *Longman Language Review*, 3, 14–17.
- Thomas, Jenny (1983). "Cross-cultural pragmatic failure". *Applied Linguistics*, 4(2), 91–112.
- Tognini-Bonelli, Elena (2001). *Corpus Linguistics at Work*. Amsterdam: John Benjamins.
- Tottie, Gunnel (1991). "Conversational style in British and American English: The case of backchannels". In K. Aijmer & B. Altenberg (Eds.), *English Corpus Linguistics* (pp. 254–271). London: Longman.
- Walsh, Steve (2002). "Construction or obstruction: Teacher talk and learner involvement in the EFL classroom". *Language Teaching Research*, 6(1), 3–23.
- Yngve, Victor (1970). "On getting a word in edgewise". In M. A. Campbell (Ed.), *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society* (pp. 567–578). Chicago: Chicago Linguistic Society.
- Zimmerman, Don & West, Candace (1975). "Sex roles, interruptions and silences in conversation". In B. Thorne & N. Henley (Eds.), *Language and Sex: Differences and Dominance* (pp. 105–129). Rowley, MA: Newbury.

Appendix

Transcription conventions for data:

A: / B:	different speakers
[]	extra linguistic information
<unintelligible>	a short indecipherable section of recorded speech
<background noise>	noise that is external to the speakers’ conversation
+	interruption
=	truncation
<\$/=> </\$/=>	unfinished (part of) sentence

Data references:

[CANCODE]	indicates that data come from Cambridge and Nottingham Corpus of Discourse in English
[LCIE]	indicates that data come from the Limerick Corpus of Irish English
[YW20a]	smaller corpus of “Young women of 20”; from the Limerick Corpus of Irish English
[YW20b]	smaller corpus of “Young women of 20”; from Cambridge and Nottingham Corpus of Discourse in English.