

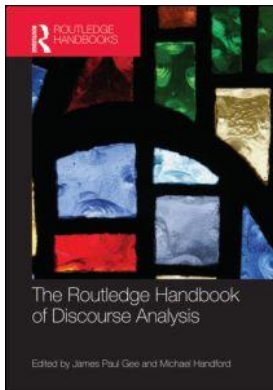
This article was downloaded by: 10.3.97.143

On: 30 Sep 2023

Access details: *subscription number*

Publisher: *Routledge*

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: 5 Howick Place, London SW1P 1WG, UK



## **The Routledge Handbook of Discourse Analysis**

James Paul Gee, Michael Handford

### **Lexis in spoken discourse**

Publication details

<https://www.routledgehandbooks.com/doi/10.4324/9780203809068.ch20>

Paula Buttery, Michael McCarthy

**Published online on: 16 Nov 2011**

**How to cite :-** Paula Buttery, Michael McCarthy. 16 Nov 2011, *Lexis in spoken discourse from: The Routledge Handbook of Discourse Analysis* Routledge

Accessed on: 30 Sep 2023

<https://www.routledgehandbooks.com/doi/10.4324/9780203809068.ch20>

**PLEASE SCROLL DOWN FOR DOCUMENT**

Full terms and conditions of use: <https://www.routledgehandbooks.com/legal-notices/terms>

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

# Lexis in spoken discourse

*Paula Buttery and Michael McCarthy*

---

## **Introduction: lexis as a discourse phenomenon**

More than a decade ago, McCarthy (1998) noted that the role of lexical patterns in written texts had been the object of detailed attention, especially within the study of lexical cohesion (Halliday and Hasan, 1976; Hasan, 1984). Similarly, the significance of multiple ties between words in written texts had been meticulously recorded by Hoey (1991). This, McCarthy asserted at the time, was not matched by anything like the same amount of research into lexical patterning in everyday spoken language. The present chapter can report some considerable progress since then, especially in light of the increased number of spoken corpus-based studies using large amounts of data and of growing interest in the study of collocation and chunking, which have in turn contributed to the methodology and findings of discourse analysis, as we demonstrate below. Chunking in particular has been examined in terms of its role in spoken interaction. In this chapter, we consider how the study of lexis using large amounts of spoken data can underpin the insights into lexical patterning already observed by keen-eyed discourse and conversation analysts in one-off extracts and can provide empirical support from a wide range of occurrences for statements concerning the regularity and recurrence of particular lexical phenomena at the level of discourse (Stubbs, 2001). The first question we address is whether there are differences between the spoken and written lexicon as a whole and what implications any differences might have for an understanding of spoken discourse. We then focus on how lexical patterns manifest themselves within and across speaker turns and their contribution to the unfolding discourse. We base our evidence on everyday, informal, spoken data, mostly social conversations, for it is there, we would argue, that patterns of negotiation and social convergence at the lexical level are most fruitfully observed. We take as uncontroversial the claim that the use of corpus data can offer considerable enhancements to discourse analysis, as demonstrated for example in Bublitz's (1988) use of corpus data in the study of cooperative conversations and, more recently, by Thornbury (2010), who argues that corpus linguistics, with its emphasis on the study of co-text, can powerfully supplement the discourse analyst's investigation of context, as well as providing large numbers of examples of given phenomena.

## **Lexis and register**

One of the many features that mark out spoken language from written language is differences in the lexicon (Lee, 2001). Such differences are commonly described as differences of register, that is to say, lexical choices are made differently depending on, in Halliday's (1978) terms, the field, mode and tenor of the situation of utterance. In face-to-face spoken interaction, interpersonal

constraints strongly influence the tenor of utterances, creating and maintaining social relations partly through lexical choices on the formal-to-informal cline (Scotton, 1985; Powell, 1992). By comparing large spoken and written corpora, it is possible quantitatively to isolate a lexicon whose probability of occurrence is much higher in, or almost exclusively confined to, spoken discourse modes, whether through the spoken medium itself, through modes that attempt to capture speech (such as fictional dialogue) or through hybrid modes such as real-time internet communications.

The lexical differences between spoken and written language may be broadly observed through a quantified comparison of word frequencies in spoken and written corpora. Distinctive items on the spoken side will then be examined in terms of their role in the creation and management of discourse. Here we consider the 2,000 most frequent items in the written (fiction) and spoken subsections of the British National Corpus (BNC, 2007). High-frequency items may be expected to form patterns at the discourse level more readily than the relatively low-frequency items that make up the rest of the English lexicon, and items that differentiate speech from writing may be analysed in terms of their roles in the creation of spoken discourse and the elaboration of social relations in conversational contexts.

We may first quantify the number of lexical items these two lists, the spoken and the written, have in common by calculating their intersection, where ‘intersection’ is calculated as twice times the number of items in both lists, divided by the number of items in the spoken list, plus the number of items in the fiction list. For these two frequency lists we find the intersection to be 0.658. This indicates that some 65 per cent of all the lexical items are found in both lists, leaving 35 per cent unique to either the spoken or written data. In Figures 20.1 and 20.2 we also consider the order of the lexical items in the frequency lists by plotting their frequency in one corpus against their frequency in the other. These graphs are plotted as log frequencies rather than raw frequencies. The reason for plotting them this way is the tendency for lexical items in a corpus to follow a Zipfian distribution. According to Zipf’s Law, the most frequent word in a corpus will generally appear about twice as often as the second most frequent, which in turn occurs twice as often as the third most frequent, and so on. A distribution such as this is difficult to plot readably on a graph, since it leads to an uneven spread of data points over a large range. On our log scale a

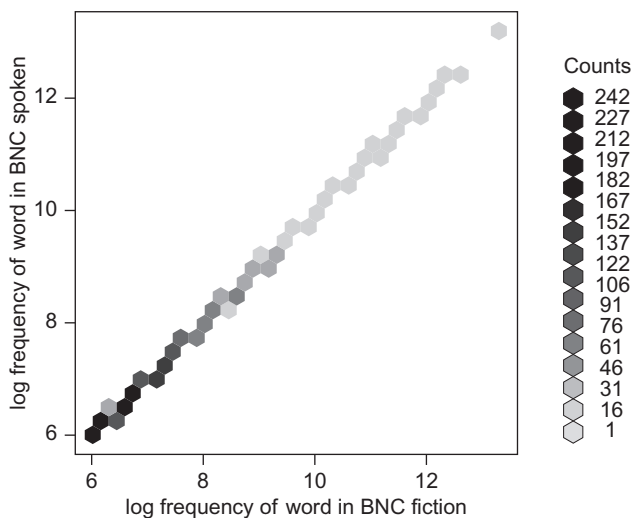


Figure 20.1 The ‘null hypothesis’ of lexical differences, spoken versus written

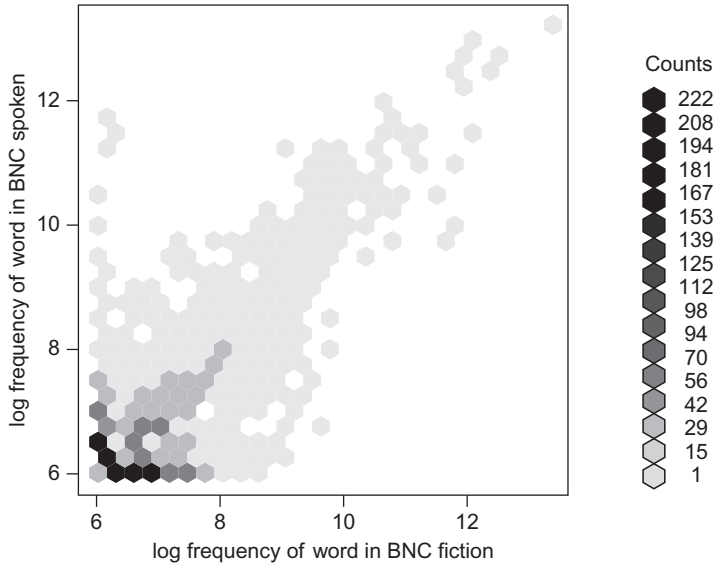


Figure 20.2 Actual variation in the data: spoken versus written frequency

lexical item occurring with a frequency of 1,000 in the written corpus and 100 in the spoken corpus would be plotted at point 3, 2 on the graph. Since there are 2,000 points to plot on each graph, hexagonal areas in Figures 20.1 and 20.2, below have been shaded to indicate the density of points in that area. A darker hexagon indicates a greater density of points plotted. Figure 20.1 shows how the data would look if the ordering of lexical items in the two lists were equivalent (the ‘null hypothesis’ graph). Figure 20.2 shows the actual data as deviating considerably from the null hypothesis. In particular we notice a column of hexagons near the y-axis. This indicates high-frequency items found in the spoken corpus that occur with substantially lower frequency in the written corpus. Otherwise the wide spread of data points from the diagonal line shows how usage in general is different between the two corpora. That is, the ordering of lexical items in the frequency lists is quite distinct.

We can further express the differences that we have illustrated graphically above as a single numeric value such as the ‘rank correlation coefficient’. This simple coefficient is calculated by considering the differences in ordering (ranking) between two lists.<sup>1</sup> The coefficient obtained will be within the range  $-1$  to  $+1$ . A value of  $-1$  indicates a complete reversal of ordering from one list to the other, whereas a value of  $+1$  indicates an identical ordering. A value of 0, meanwhile, indicates that the two orderings are independent of each other. For our frequency lists, the rank correlation coefficient is approximately 0.5. This outcome indicates that, while on the whole the rankings are similar, there is some degree of distinctiveness between the two lists. We should note that rank correlation is a simple metric that treats all ranks with equal importance. Perhaps a more appropriate metric for long lists of lexical items would be one that punishes more heavily deviation between highly ranked items. The problem goes back to the Zipfian distribution described above: there will be many low-frequency items at the bottom of the list, which will tie for ranking. However, for shorter lists like ours, the rank correlation gives a good idea of important deviations of ordering.

What, then, is the nature of these differences between the spoken and written lexicons and what functions do the spoken elements carry out at the discourse level? McCarthy and Carter

(1997) showed how the top 50 most frequent words in a spoken corpus differed from the top 50 in a written corpus of the same size, and how discourse-marking items such as *know* (in its high-frequency chunked manifestation of *you know*—see below, section ‘lexical items in discourse: chunks’), *right* and *well* featured in the spoken list but did not appear in the top 50 written list, giving further substance to the attention paid to these items by discourse analysts over a number of years. Equally, the very high frequency of items such as *just* and *think* (most commonly chunked as *I think*) in the spoken list pointed to the interpersonal strategies of hedging and politeness, which are central to cooperative and successful face-to-face interaction. Interpersonal exigencies seemed, therefore, to run high in the spoken lexicon. In other words, the items that are of unusually high frequency in the spoken corpus are implicated in the creation and maintenance of social relations and successful interaction; their roles are not merely semantic, but rather located in the pragmatics of discourse.

Further down our BNC frequency lists, many lexical items have overwhelmingly more occurrences in spoken discourse than in written discourse, and vice-versa. Certain morphological types cluster on one list or the other. For example, evaluative adjectives ending in *-y* such as *yucky*, *stropy*, *comfy* and *grumpy* have greatly differing distributions in the spoken and written components of the BNC (Figure 20.3).

Conversely, in the same corpora, a set of nouns indicating facial expressions (*grimace*, *scowl*, *smirk* and *pout*) are relatively rare in spoken discourse (Figure 20.4).

These different distributions are not absolutes; they are simply high probabilities which impart to spoken and written discourses their different lexical fingerprints and, in their institutionalized uses, have direct implications for the kinds of relationships projected and created among interlocutors.

But quantitative differences of register such as frequencies in speech versus writing are not enough to account for the contribution of lexis to discourse, albeit they may reveal something about the nature of informality in relationships among participants in banal, everyday conversations. Of greater interest to the discourse analyst is how lexical items are used in real contexts, so we now turn first to the discourse roles of lexical chunks (whose identification derives from quantitative measures) and then to how the use of lexical items unfolds in the developing discourse and to an exegesis of the functions of lexical choice within and across speaker turns. We still have recourse to our corpus of data, but our preoccupation is now with the local rather than the global, with the qualitative rather than the quantitative interpretation of spoken data.

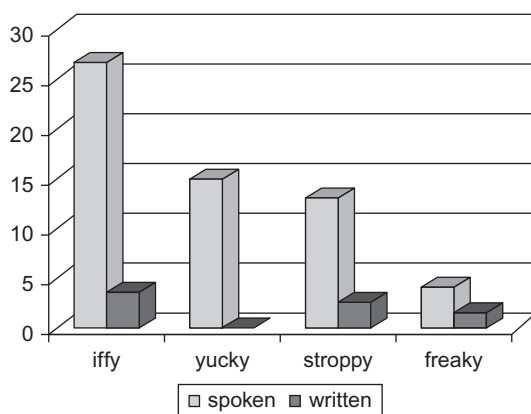


Figure 20.3 Frequency of *-y* adjectives (BNC) per 10m words

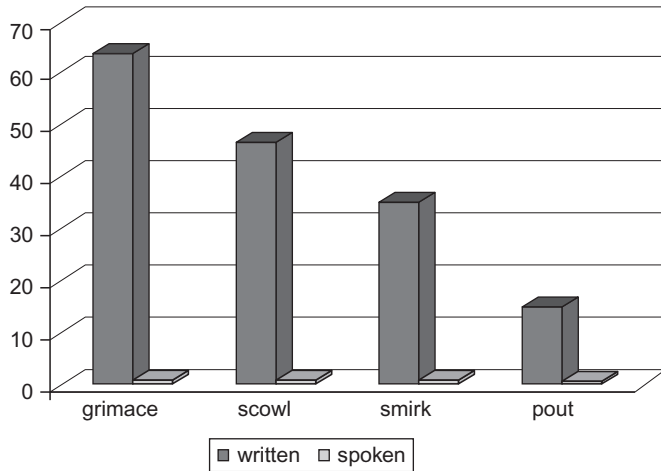


Figure 20.4 Frequency of facial expression nouns (BNC) per 10 m words

### Lexical items in discourse: chunks

In this chapter we consider the discourse roles not only of single words, but of lexical chunks. The idea that strings of words may have integrated meanings and that such strings may develop specialized functions is not a new one (see Bolinger, 1976; Pawley and Syder, 1983). More recently, Sinclair (1987a, b, 1991) has proposed two basic principles that operate in the enactment of meaning. The first of these, the open choice principle, is the conventional notion that grammatical slots are ‘filled’ by lexical items. The second and more radical one is the *idiom principle*, which asserts that language users dispose of a huge repertoire of lexicogrammatical chunks whose form and meaning are pre-established and in a delicate relationship with each other. Chunks, as we shall refer to them in this chapter, have been extensively studied under different names, but with similar emphasis and preoccupations. Terms for the phenomenon, apart from widely used labels such as *fixed expressions* and *multi-word units*, include *routine formulae* (Coulmas, 1979), *lexicalized stems* (Pawley and Syder, 1983), *formulaic sequences* (Wray, 2002; Schmitt, 2004), *chunks* (O’Keeffe *et al.*, 2007) and *lexical bundles* (Biber and Conrad, 1999).

The significance of lexical chunks in spoken discourse is evidenced by the high frequency of items with interactive meanings, which are often as frequent as, or more frequent than, common, everyday single words. The most frequently occurring chunks in the five-million-word CANCODE<sup>2</sup> spoken corpus (for details of the corpus, see McCarthy, 1998) include items such as *you know*, *you know what I mean*, *(and) (all) that sort of thing*, *and all the rest of it* and various other items which project assumptions on the speaker’s part and invite the listener to converge (see below). In this respect they play an important part in the creation of adjacency sequences; that is to say, for conversation to proceed normally and coherently, they invite listeners to respond as soon as possible after the utterance of the items (hence ‘adjacency’) (see also Chapter 9 in this volume). The high-frequency chunks also include discourse management items such as *so anyway*, which occurs more than 120 times in the CANCODE spoken corpus and is used to mark discourse boundaries, often a return to a series of narrative events after a descriptive or evaluative segment or an aside, or a shift to a new (sub-)topic. Extract 1 illustrates the narrative function.

1

[speaker is talking about making arrangements to view a house for a possible shared rental with friends]

... they rang me up, and I said look, you know, I've got, I've got my exams today and I really don't want to come up unless you see something, if you do, I will come up because they reckon it's a joint effort, if we're gonna live together, we've all got to see it and make a contribution to looking at the house. **So anyway**, they rang me up Friday night, and said, oh, seen this house, it's lovely and really nice. What's the arrangement? We'll meet you there at twelve o'clock Saturday, right. Right, I'll meet you at McDonald's blah blah blah. **So anyway**, we got down here,...

(BNC)<sup>3</sup>

Other chunks that have conversational-management functions include *talking of (x)*, *now you come to mention it*, *right then*, *now then*, *the thing is*, *there you go*, etc. These chunks may also be discontinuous, in that they may offer a slot within them that can be filled by an open choice. Such a chunk may be exemplified by *the x thing is*, where *x* is typically realized by an evaluative adjective phrase (frequent *x*-fillers in the spoken corpus include *best*, *worst*, *funny*, *silly*), or else an information-focus phrase such as *the (most) important thing is*, *the only thing is*, *the first thing is*, *the other thing is*. Extracts 2 and 3 exemplify this group.

2

<\$1> So if you go to London really **the best thing is** if you can get a research job and a contract and reasonable accommodation for six months.

<\$2> Er yeah. (CANCODE © Cambridge University Press)

3

<\$1> Cos it 's such er a light room.

<\$2> Mm.

<\$1> **The only thing is** with that radiator it still, it does get very cold in here. Because of all the glass. (BNC)

These lexical chunks have developed pragmatic specialisms in discourse with regard to organization and management and the signalling of stance, and operate outside of clause- and sentence-boundaries, as free-standing discourse items. *The only thing is* typically functions as a signal of a situation that may need attention or is problematic in some way. Such chunks function as 'long words' and co-exist with the many single words that have been recognized as having discourse-organizing or stance-signalling functions (e.g. *well*, *right*, *so*, *wow*, *anyway*, *absolutely*, *fine*, etc).

## Lexical repetition and relexicalization

One of the most immediate and visible features of lexis beyond the sentence level in spoken interaction is the repetition of lexical items, which occurs both within and across speakers' turns. Persson (1974) used spoken and written data to examine repetition and included an examination of repeated modifiers and the friendly repetition of greetings and farewells. Persson's definition of repetition restricted itself to sequential repetition, that is to say, to the immediate repetition of identical lexical items by one speaker or writer (p. 11). Later work has stressed the importance of looking across turn boundaries to how repetition of lexis occurs between speakers in conversation. Bublitz (1988) used corpus data from the Survey of English Usage to look at a wide range of phenomena that manifest across speaker-boundaries in the creation and maintenance of supportive and cooperative conversations, within which he observed a number of examples of direct

lexical repetition. In Bublitz's examples (p. 229), second speakers repeat one or more words used by first speakers to support the first speaker's utterance, further examples of which we provide below. Lindström (2001) (using Swedish data, though the same applies to English) notes how everyday greetings and other adjacency pair routines exploit lexical repetition (e.g. *Good morning – Good morning*) and comments on the range of interpretations that can be put on such examples, including the iconic nature of reduplication (see, for example, the type of phenomena attested in Wang, 2005) and considerations of politeness – for instance, how reduplicated routines may reinforce positive politeness (Brown and Levinson, 1987). Lindström (2001) also notes how reduplication in adjacency pairs can mark discourse boundaries; examples in English would include a second speaker repeating *right* or *fine* to mark a conversational (pre-)closure. Meanwhile, Carter (2004: 6–8) comments upon the role of repetition in everyday creativity in speaking (see also Chapter 22 in this volume).

Conversation analysis (CA) research has shed light on the functions of repetitions and on their place in sequences of turns. Schegloff (1997) argues that repetitions in conversation can raise the need for repair, can register receipt or can target a next action (see also Kim, 2002). A number of other studies have also taken a CA approach to lexical repetition and adjacency, explicating repetition in terms of focusing on items in a speaker's turn in order to acknowledge receipt and/or understanding, or to highlight a problem in the hearing or understanding of a highlighted item by a second speaker, to enact some organizational move in the talk, or to project stance (Wong, 2000; Svennevig, 2004; Koshik, 2005: ch. 5; Stivers, 2005). Tannen (1989) is an important study of repetition in which she examines a range of phenomena, including exact repetition and reformulation of the same ideas in different words, a feature we exemplify below.

McCarthy and Carter carried out research into patterns of lexis in spoken data, focusing on repetition and relexicalization (the repetition of content in modified lexical form – see below) within and across speaker turns in conversation (McCarthy, 1988, 1992; McCarthy and Carter, 1994: ch. 3). Their work was concerned with how speakers use repetition to create and sustain convergence or, on occasion, to signal resistance or divergence. They examined phenomena in conversation that included direct repetition, as in extracts 4 and 5:

- 4  
 <\$1> Yeah it's **nice** isn't it.  
 <\$2> Yeah. It is **nice**.  
 <\$1> **Nice nice** present. (CANCODE © Cambridge University Press)
- 5  
 <\$2> [Road name] Street's got so many houses on it that people won't move into even the homeless won't move into them because of the **crime**.  
 <\$3> **Crime problem**.  
 <\$2> **Crime problem**. But I mean whether that's exaggerated or not I don't know.  
 (CANCODE © Cambridge University Press)

In extracts 4 and 5, speakers converge and create or maintain solidarity through repeating each other's words. In the case of extract 4, the sequence pivots around the repetition of *nice* across three turns. In extract 5, the first speaker's use of *crime* is incremented by the second speaker's adding *problem*, a collocation which is then taken up and repeated by the first speaker. Sometimes, however, the repetition of an entity may be realized by a relexicalized form, typically a (near-)synonym, as in extract 6, where the same speaker presents *fellow* and *chap* as synonymous.



6

... but he 's a very nice **fellow** now, very nice **chap**. (BNC)

A similar kind of solidarity to that in extracts 4 and 5 also seems to be achieved by rewording with near-synonymous expressions by a second or further speaker, as in extracts 7 and 8.

7

<\$3> I think he's a **brilliant** actor.

<\$2> Yeah. He is **very good**.

<\$3> He's a **really great** actor.

<\$2> Yeah. (CANCODE © Cambridge University Press)

8

<\$2> What what a **beautiful day**.

<\$4> It's **lovely** isn't it.

<\$2> Oh.

<\$5> Yeah. It's **fantastic**.

<\$1> A **great day** for the match. (CANCODE © Cambridge University Press)

Such patterns of relexicalization are commonplace in informal conversation. They serve a purpose in terms of content, in that they confirm understandings through negotiation of meanings, that is to say the local, contextual meaning of items is an achievement, arrived at through the proffering of terms by different speakers who converge with (extract 9) or renegotiate (extract 10) the use of particular items by their interlocutors.

9

[Speaker 1 addresses a cat, who has just greedily devoured a piece of chicken; the other speakers comment on the event]

<\$1> God have you eaten that already?

<\$2> Yeah.

<\$1> I'm surprised he didn't choke.

<\$2> I wouldn't say he **wolfed** it but he **lioned** it.

<\$1> **Catted** it.

<\$2> He **lioned** it.

<\$1> Yeah.

<\$2> **Tigered** it.

<\$1> Or something like that. (CANCODE © Cambridge University Press)

10

[discussing speaker <\$1>'s new job]

<\$1> I've moved on and dealing with different things now.

<\$2> Mm.

<\$1> So erm-

<\$2> So you left us for something **better**. [laughs]

<\$1> **I wouldn't say better**. It's **different**. And I think you learn **different** things in each place that you go to. (CANCODE © Cambridge University Press)

In this way the establishment of lexico-semantic relations may be seen as interactive and, in the case of extract 9, creative (see Carter, this volume). The patterns also serve a relational purpose in that they create and consolidate social understandings and compacts. There are, of course, many occasions where speakers cannot or do not wish to converge lexically. In most such cases, in

non-conflictual conversations, disagreements are hedged and counter-evaluations are downtoned rather than stated baldly, as in extracts 10 (above) and 11.

11

<\$3> Corfu is **awful** is it?

<\$2> Er **not really** no parts of it are **okay**. (CANCODE © Cambridge University Press)

Although syntax and phonology (see Ogden, 2006) clearly play a role in the construction of responsive turns, the lexical contribution to the preference organization of adjacency pairs is apparent, not just in formulaic sequences, but in the repetition and relexicalization we have exemplified here – a phenomenon underlined in the work of CA analysts such as Pomerantz (1984).

One aspect of repetition and relexicalization is manifested in the common practice of listeners to predict what a speaker is about to say, to utter the predicted words and then to receive confirmation of their prediction either through exact repetition or near-repetition or in the form of a relexicalization. By searching in a corpus for symbols denoting latched turns (typically a + sign), it is possible to observe the phenomenon as it unfolds. In a latched turn, a speaker is interrupted and then resumes his/her incomplete turn. The interruption is usually a cooperative act, perhaps designed to supply a word where there is evidence of hesitation, as in extract 10.

10

[<\$?F> = an unidentifiable female speaker]

<\$6> What was it in the little box?

<\$?F> Oh I dunno.

<\$2> Oh those little erm+

<\$1> The **charms**.

<\$2> +those little **charms**. (CANCODE © Cambridge University Press)

The latching may be more of an overlap than an interruption, providing an occasion for the almost simultaneous use of the same word(s), as in extract 11.

11

<\$1> Well quite honestly I said I despair coming down off the bus I thought well if this is the **future** | **generation**+

<\$2> | **Future generation**. Yeah.

<\$1> +God help us. (CANCODE © Cambridge University Press)

## Lexical triggers

Extracts 4–8 above consisted of short turns wherein an evaluative lexical item is taken up by another speaker and repeated or recast in a subsequent turn. The fact that the repeated or recast items are evaluative is important, in that evaluative claims lay the speaker open to contradiction or challenge, though by its very nature collaborative, non-conflictual conversation will tend towards supportive responses and, as we have seen, counter-evaluations may be downtoned or downgraded (Pomerantz, 1984).

In conversational terms, evaluative claims are likely to trigger speaker change, in order for the subsequent speaker to be able to support (or challenge) the claim. Evison and McCarthy (in press) investigated this hypothesis in a 1-million-word sub-corpus of social and intimate conversations in the CANCODE corpus and found that a large proportion of utterances containing evaluations using the most frequent evaluative adjectives did indeed trigger speaker change. Evison and McCarthy found that, for the most frequent evaluative adjectives (which included, for example,

*good, awful, true, nice, funny, lovely*), the percentages of the total occurrences that immediately preceded a listener response ranged from the lower 20s in the case of *good, bad* and *great*, to 34–37 per cent for *lovely, awful* and *horrible*, right up to 45 per cent and 49 per cent in the case of *brilliant* and *true*, respectively. Responses on occasion consist of no more than backchannels such as *mm* or *uhuh*, but they may also consist of more extended reactions. Examples of this phenomenon of lexical triggering abound in the CANCODE corpus and reveal another aspect of the role of lexis in the construction of discourse.

12

- <\$1> I don't always agree the customer's always right I don't think that's **true**.  
 <\$1> No.  
 <\$2> But you've got to create an environment such as service level agreements, er monitoring meetings things like that.  
 <\$1> Yeah.  
 <\$2> Where you can have a dialogue with the customer. (CANCODE © Cambridge University Press)

13

- <\$1> Right let's get on to the positive one then, your holiday.  
 <\$2> Yes it was **brilliant**.  
 <\$1> **Good**.  
 <\$2> We went on the fourth of June er for a fortnight and it was absolutely wonderful. (CANCODE © Cambridge University Press)

Evison and McCarthy (in press) also observed that vague language items such as *(and) things like that, or whatever, and so on, (and) that sort/kind of thing* regularly triggered speaker change – understandably, given the projection of shared knowledge that such items encode, inviting the listener(s) to concur with the assumed content, as in extracts 14 and 15.

14

- <\$2> We also monitor complaints.  
 <\$1> Is this through Jim's | department?  
 <\$2> | Jim yeah. Yeah. And we also monitor er er capture er letters of praise and thank-yous **and things like that**.  
 <\$1> **Bouquets**.  
 <\$2> **Er yeah**. Em and of course we also have er the er Community Health Council drop in on us from time to time. (CANCODE © Cambridge University Press)

15

- <\$5> You don't get private estates around factories **or whatever**+  
 <\$1> **Right**.  
 <\$5> +chemical plants, not in my experience. (CANCODE © Cambridge University Press)

It may also be noted that extract 12 above contained a further example of this phenomenon (*monitoring meetings things like that*).

### Lexis and turn-openings

What we have attempted to show so far is the degree to which lexis is implicated in the successful construction of sequences of speaker turns, and we focused on how lexis is (a) taken up in

subsequent turns and (b) may be a trigger for speaker change. A third dimension of turn construction is turn-opening. Tao (2003) investigated turn-openings in a spoken corpus and found that turn-initial items are mostly lexical and syntactically free-standing. Tao found in the turn-initial slot a high incidence of items such as *yes, well, right, okay* and of chunks such as *I think, you know, I mean, that's + adjective*. Tao's work underscores the way speakers attend to the prior turn and to relational concerns before they launch into their own message. Turn-initial items convey interactional meaning and maintain the flow of the talk by linking back to what has just been said. McCarthy (2010: 7–8), reinforcing Tao's work, shows that the top 20 turn-openers in the CANCODE corpus include common responsive items such as *yeah, yes, no, oh*, along with linking words such as *and, but* and *so*, backchannel items such as *mm*, and discourse-organizing words such as *well* and *right*. By far the most frequent word in a 5-million-word mixed written corpus, the definite article *the* occurs as sentence-initial in more than 23,000 written sentences (out of a total of 320,000, i.e. just over 7 per cent), yet only occupies the initial position in 4,300 of some 478,000 speaker turns (less than 1 per cent) in the 5 million words of the CANCODE corpus. Over 130 of these consist of chunks based around *the* such as *the thing is, the trouble is, the only thing is*, etc., which we examined above in section on lexis and register. The reason for dwelling on these numbers is that they underscore the lexical nature of turn openers and the way the free-standing lexical turn-openers 'push down' even such common words as the definite article, which might be expected to appear in almost any position in the speaker turn. In fact second, third, fourth and fifth position in the turn totals for the definite and indefinite articles all individually well exceed initial position totals for each of the articles. In other words, grammatical items such as the articles are dispreferred at the beginning of speakers' turns, in favour of the free-standing lexical turn-openers.

### Response tokens as discourse items

As we have seen in the extracts presented so far, short responses are frequent in spoken discourse and may range from backchannels such as *mm* and *uhum*, through reactives such as *oh*, variations on *yes* and *no*, to fully lexical items such as *right* and *good*. This latter group of items, which in their grammatical identity belong mostly to the word class of adjectives and adverbs, includes single words such as *good, right, fine, lovely, wonderful, cool, marvellous, great, excellent, true, absolutely, definitely, certainly, exactly* – either used alone or along with *yes, yeah, no, okay* or with *that's*, as in extracts 16–18.

16

[talking about VAT, value added tax, a tax on goods and services]

<\$1> Really V A T's, I think, and I think it should only be paid when it, when the invoices+

<\$2> **Exactly!**

<\$1> +have been cleared, rather than always paying+

<\$2> **Exactly!**

<\$1> +for that quarter+

<\$2> Yeah.

<\$1> +even if you 've not got paid.

<\$2> Yeah. **Exactly!** (BNC)

17

[talking about someone who accepted an acting job]

<\$1> The thing is she couldn't have turned it down because she hadn't done any work for two years so she had to.

<\$2> **Oh yeah absolutely.**

<\$1> And she got to film on location in Majorca. (CANCODE © Cambridge University Press)

18

<\$1> Okay. Erm er you just want me to send these on presumably.

<\$2> Yeah. If you wouldn't mind.

<\$1> **That's fine.**

<\$2> Okay. Thanks.

<\$1> Okay then.

<\$2> Bye.

<\$1> Bye. (CANCODE © Cambridge University Press)

Some of these items are pragmatically specialized, for example the use of *fine* in (pre-)closing sequences (extract 18) or of *certainly* in response to a request, or of *right* to mark major discourse boundaries in longer events such as meetings and lessons (see Sinclair and Coulthard, 1975). The items may also be reduplicated, and it is not uncommon to find them repeated several times in one turn, expressing enthusiasm or extra engagement, as in extract 19.

19

[at an estate agent's: the customer is seeking a property in the stated locations]

<\$1> Two bedroomed, no? Round Arford, Johnston and Milford

<\$2> Yeah we 've got quite a few actually.

<\$1> Have you?

<\$2> Mm

<\$1> **Good good good good good.** Thank you. (BNC)

McCarthy (2002, 2003) calls such items 'non-minimal response tokens' (in the sense that they do more than the minimum requirement of acknowledgement or saying yes or no) and presents evidence of their widespread use. He attributes to them the function of displaying 'good listener-ship', in that they do more in both propositional and relational terms than bare *yes* or *no* and can display a high degree of involvement and interactivity without grabbing the floor and taking over the main speaker's role. Such items have almost shed their word-class identity as adjectives or adverbs and merit a separate classification as discourse items, displaying the full lexicality of discourse markers and free-standing, beyond-the-clause capabilities. Fries (1952: 49) long ago noted the responsive function of items such as *good*, while Duncan (1974) expanded the notion of backchannel (e.g. vocalizations such as *mm*) to include items such as *right* and *I see*. Pomerantz (1984), as we have noted, focused on the evaluative force of responses in second-pair parts of adjacency pairs. Öreström (1983) also broadened the purview of backchannel responses to examine fully lexical response items such as *quite* and *good*. Similarly, Tottie (1991) examines a range of responsive actions from body language, through vocalization, through single word responses, through phrasal/chunked responses, to short clauses and longer utterances, covering the whole cline to the point where a listener unequivocally takes over the speaker role. For our present purposes, though, it is the lexical nature of response tokens that is at the heart of their discourse roles.

### Lexical chaining within and across speaker boundaries

In the section on lexical items in discourse (chunks), we considered repetition and relexicalization in relation to their role in the creation of adjacency sequences and interactional convergence.

However, uses of (near-)synonyms and relexicalized forms can carry over longer stretches of discourse to create lexical chains that allow us to follow both topical development and speakers' changing stances as they converge and occasionally diverge. Such chains may be developed both within an individual speaker's turn and across turn boundaries. In the examples below, the items contributing to chains are in bold. In example 20, the same speaker offers *fantastic*, *brilliant*, *nice* (x2) and boosted *really nice* to comment on a holiday just taken.

- 20
- <\$2> Ah well I went on holiday as you know.
- <\$1> Mhm.
- <\$2> That was **fantastic** that was. Oh dear it was **brilliant**. I didn't wanna come back at all. Oh it was **nice**. Mm. Beautiful sunshine beautiful beaches. Oh everything was **nice**.
- <\$1> Mm.
- <\$2> It was **really nice**.
- <\$1> Good. (CANCODE © Cambridge University Press)

Such relexicalizations over several turns are necessary inasmuch as exact repetition of the initial evaluation *fantastic* four times in a row would be pragmatically odd and stylistically untypical (not to mention quite tedious for the interlocutors!). Where more than one speaker takes up an item, exact repetition seems more naturally allowable, as in extract 21.

- 21
- [In a shoe-shop: <\$1> is the customer, <\$2> is the assistant]
- <\$2> Probably needs adjusting but I'll check that.
- <\$1> Oh right. That's **lovely**.
- <\$2> | Okay.
- <\$1> | Yeah | that's **nice**.
- <\$2> | They're **nice** aren't they.
- <\$1> Yeah they are **nice**.
- <\$2> **Very very nice**.
- <\$1> Thank you.
- <\$2> They feel **right**?
- <\$1> Yeah.
- <\$2> Does it?
- <\$1> That feels **pretty good** actually.
- <\$2> Yes **smashing fit**.
- <\$1> Yeah. (CANCODE © Cambridge University Press)

Extract 21 looks, on the face of it, to be a special and different case (a service encounter rather than a social conversation), but it underlines how social convergence can be exploited by servers and their clients to create good trading relations at the interpersonal level (see McCarthy (2000) for further examples).

## 9. Conclusion

Lexical patterning at the discourse level may be seen to be an important feature of spoken interaction. The choice of registers within the lexicon has important implications for the creation and maintenance of particular types of social relations and is a resource to be exploited or manipulated (a feature also noted and discussed by critical discourse analysts; see for example, Bhatia, 2006; Patrona, 2006). Beyond register, lexical patterns in spoken discourse are evidenced

not only in the way speakers negotiate meanings through the trading and uptake of lexical items, but also in the ways social relations are projected, created and maintained through the sharing of lexical items and through the constant weaving and re-weaving of lexical meanings within and across turns. Indeed, we have attempted to show the centrality of lexis in turn-construction, in the sense of its role in effectively linking one turn to another and also in its function of responsiveness to incoming talk and in the part it plays in maintaining conversational flow or 'confluence' (McCarthy, 2010: 7–8) and in the projection of listenership.

## Further reading

McCarthy, M. J. (1998) *Spoken Language and Applied Linguistics*. Cambridge: Cambridge University Press.

In this book McCarthy devotes a chapter to lexical patterning in spoken language, summing up work to date carried out in collaboration with Ronald Carter and others and adding evidence from the 5-million-word CANCODE corpus of everyday spoken English. The chapter shows corpus examples of repetition, relexicalization, negotiation of topic and listeners' contributions to lexical patterns in spoken discourse.

O'Keeffe, A., McCarthy, M. J. and Carter, R. A. (2007) *From Corpus to Classroom*. Cambridge: Cambridge University Press.

This book has several chapters in which lexical aspects of spoken language are dealt with. The authors use corpus evidence to illustrate the ubiquity of lexical chunks of spoken discourse as well as looking at the occurrence and functions of idiomatic expressions, the interrelationship between lexis and grammar, the role of listeners, hedging, vagueness and the use of discourse markers in everyday spoken language. The volume also includes sections on creativity in everyday discourse and special examples of spoken discourse such as academic talk and second-language classroom data.

Powell, M. J. (1992) 'Semantic/pragmatic regularities in informal lexis: British speakers in spontaneous conversational settings', *Text* 12 (1): 19–58.

Using the London–Lund spoken corpus, this study investigates the distribution and use of informal lexis in a database of conversations. An inventory of lexical items is drawn up and refined, which leads to the categorization of lexical items and their functions in spoken discourse. The principal functions discussed are evaluative and expressive ones, and idiomatic expressions, vague language and intensification are considered among and across different speakers.

## Notes

- 1 For this calculation it is necessary that the compared orderings contain exactly the same set of items. The calculation was essentially carried out on the 65% intersection of the items that occur in both frequency lists.
- 2 CANCODE means Cambridge and Nottingham Corpus of Discourse in English. The corpus consists of five million words of informal conversations recorded across the islands of Britain and Ireland. Cambridge University Press is the sole copyright holder.
- 3 Where indicated, the data cited herein have been extracted from the British National Corpus Online service, managed by Oxford University Computing Services on behalf of the BNC Consortium. All rights in the texts cited are reserved.

## References

- Bhatia, A. (2006) 'Critical discourse analysis of political press conferences', *Discourse and Society*, 17 (2): 173–203.
- Biber, D. and Conrad, S. (1999) 'Lexical bundles in conversation and academic prose', in H. Hasselgard and Oksefjell, S. (eds.) *Out of Corpora: Studies in Honor of Stig Johansson*. Amsterdam: Rodopi, pp. 181–190.
- Bolinger, D. (1976) 'Meaning and memory', *Forum Linguisticum*, 1: 1–14.
- Brown, P. and Levinson, S. (1987) *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.
- Bublitz, W. (1988) *Supportive Fellow-Speakers and Cooperative Conversations*. Amsterdam: John Benjamins.

- Carter, R. A. (2004) *Language and Creativity: The Art of Common Talk*. London: Routledge.
- Coulmas, F. (1979) 'On the sociolinguistic relevance of routine formulae', *Journal of Pragmatics*, 3: 239–266.
- Duncan, S. (1974) 'On the structure of speaker-auditor interaction during speaker turns', *Language in Society*, 2: 161–180.
- Evison, J. and McCarthy, M. J. (in press) 'Social talk', in A. Barron and K. Schneider (eds.) *Pragmatics of Discourse*. Berlin: Mouton de Gruyter.
- Fries, C. C. (1952) *The Structure of English*. New York: Harcourt, Brace.
- Halliday, M. A. K. (1978) *Language as Social Semiotic*. London: Edward Arnold.
- Halliday, M. A. K., and Hasan, R. (1976) *Cohesion in English*. London: Longman.
- Hasan, R. (1984) 'Coherence and cohesive harmony', in J. Flood (ed.) *Understanding Reading Comprehension*. Newark, Delaware: International Reading Association, pp. 181–219.
- Hoey, M. P. (1991) *Patterns of Lexis in Text*. Oxford: Oxford University Press.
- Kim, H. (2002) 'The form and function of next-turn repetition in English conversation', *Language Research*, 38 (1): 51–81.
- Koshik, I. (2005) *Beyond Rhetorical Questions: Assertive Questions in Everyday Interaction*. Amsterdam/Philadelphia, PA: John Benjamins.
- Lee, D. Y. W. (2001) 'Defining core vocabulary and tracking its distribution across spoken and written genres: evidence of a gradience of variation from the British national corpus', *Journal of English Linguistics*, 29: 250–278.
- Lindström, J. (2001) 'Testing Östman: iconicity and reduplicative politeness routines in Swedish', in V. Laakso, J. Leino, and J. Raukko (eds.) *Postcards, Implicitness, Constructions: Festschrift to Professor Jan-Ola Östman*. Vantaa: Tummavuoren kirjapaino, pp. 276–284.
- McCarthy, M. J. (1988) 'Some vocabulary patterns in conversation', in R. A. Carter and M. J. McCarthy (eds.) *Vocabulary and Language Teaching*. London: Longman, pp. 181–200.
- McCarthy, M. J. (1992) 'Interactive lexis: prominence and paradigms', in R. M. Coulthard (ed.) *Advances in Spoken Discourse Analysis*. London: Routledge, pp. 197–208.
- McCarthy, M. J. (1998) *Spoken Language and Applied Linguistics*. Cambridge: Cambridge University Press.
- McCarthy, M. J. (2000) 'Captive audiences: the discourse of close contact service encounters', in Coupland, J. (ed.) *Small Talk*. London: Longman, pp. 84–109.
- McCarthy, M. J. (2002) 'Good listenership made plain: British and American non-minimal response tokens in everyday conversation', in R. Reppen, S. Fitzmaurice, and D. Biber (eds.) *Using Corpora to Explore Linguistics Variation*. Amsterdam: John Benjamins, pp. 49–71.
- McCarthy, M. J. (2003) 'Talking back: "small" interactional response tokens in everyday conversation', in J. Coupland (ed.) Special issue of *Research on Language and Social Interaction* on 'Small Talk'. 36 (1): 33–63.
- McCarthy, M. J. (2010) 'Spoken fluency revisited', *English Profile Journal*, 1 (1): 1–15. Available at <http://journals.cambridge.org/action/displayJournal?jid=EPJ> (accessed 1 March 2011).
- McCarthy, M. J. and Carter, R. A. (1994) *Language as Discourse: Perspectives for Language Teaching*. London: Longman.
- McCarthy, M. J. and Carter, R. A. (1997) 'Written and spoken vocabulary', in N. Schmitt and M. J. McCarthy (eds.) *Vocabulary: Description, Acquisition, Pedagogy*. Cambridge: Cambridge University Press, pp. 20–39.
- Ogden, R. (2006) 'Phonetics and social action in agreements and disagreements', *Journal of Pragmatics*, 38: 1752–1775.
- O'Keefe, A., McCarthy, M. J., and Carter, R. A. (2007) *From Corpus to Classroom*. Cambridge: Cambridge University Press.
- Öreström, B. (1983) *Turn-Taking in English Conversation*. Lund: Gleerup.
- Patrona, M. (2006) 'Conversationalization and media empowerment in Greek television discussion programs', *Discourse and Society*, 17 (1): 5–27.
- Pawley, A. and Syder, F. (1983) 'Two puzzles for linguistic theory: nativelike selection and nativelike fluency', in J. Richards and R. Schmidt (eds.) *Language and Communication*. New York: Longman, pp. 191–226.
- Persson, G. (1974) *Repetition in English: Part I, Sequential Repetition*. Uppsala: Acta Universitatis Upsaliensis.
- Pomerantz, A. (1984) 'Agreeing and disagreeing with assessments: some features of preferred/dispreferred turn shapes', in J. Atkinson, and J. Heritage, (eds.) *Structures of Social Action*. Cambridge: Cambridge University Press, 57–101.
- Powell, M. J. (1992) 'Semantic/pragmatic regularities in informal lexis: British speakers in spontaneous conversational settings', *Text*, 12 (1): 19–58.



- Schegloff, E. (1997) 'Practices and actions: boundary cases of other-initiated repair', *Discourse Processes*, 23 (3): 499–545.
- Schmitt, N. (ed.) (2004) *Formulaic Sequences*. Amsterdam: John Benjamins.
- Scotton, C. (1985) 'What the heck, sir: style shifting and lexical colouring as features of powerful language', in R. Street, J. Capella, and H. Giles (eds.) *Sequence and Patterning in Communicative Behaviour*. London: Edward Arnold, pp. 103–119.
- Sinclair, J. McH. (1987a) 'The nature of the evidence', in J. McH. Sinclair (ed.) *Looking up*. Glasgow: Collins, pp. 150–159.
- Sinclair, J. McH. (1987b) 'Collocation: a progress report', in R. Steele, and T. Threadgold (eds.) *Language Topics: An International Collection of Papers by Colleagues, Students and Admirers of Professor Michael Halliday to Honour Him on His Retirement*. Amsterdam: John Benjamins, Volume 2, pp. 319–331.
- Sinclair, J. McH. (1991) *Corpus, Concordance and Collocation*. Oxford: Oxford University Press.
- Sinclair, J. McH. and Coulthard, R. M. (1975) *Towards an Analysis of Discourse*. Oxford: Oxford University Press.
- Stivers, T. (2005) 'Modified repeats: one method for asserting primary rights from second position', *Research on Language and Social Interaction*, 38 (2): 131–158.
- Stubbs, M. (2001) *Words and Phrases: Corpus Studies in Lexical Semantics*. Oxford: Blackwell.
- Svennevig, J. (2004) 'Other-repetition as display of hearing, understanding and emotional stance', *Discourse Studies*, 6 (4): 489–516.
- Tannen, D. (1989) *Talking Voices*. Cambridge: Cambridge University Press.
- Tao, H. (2003) 'Turn initiators in spoken English: a corpus-based approach to interaction and grammar', in P. Leistyna and C. F. Meyer (eds.) *Corpus Analysis: Language Structure and Language Use*. Amsterdam: Rodopi, pp. 187–207.
- The British National Corpus*, version 3 (BNC XML Edition) (2007) Distributed by Oxford University Computing Services on behalf of the BNC Consortium. Available at <http://www.natcorp.ox.ac.uk/>.
- Thornbury, S. (2010) 'What can a corpus tell us about discourse?', in A. O'Keeffe and M. J. McCarthy (eds.) *The Routledge Handbook of Corpus Linguistics*. Abingdon: Routledge, pp. 270–287.
- Tottie, G. (1991) 'Conversational style in British and American English, the case of backchannels', in K. Aijmer and B. Altenberg (eds.) *English Corpus Linguistics*. London: Longman, pp. 254–271.
- Wang, S.-P. (2005) 'Corpus-based approaches and discourse analysis in relation to reduplication and repetition', *Journal of Pragmatics*, 37 (4): 505–540.
- Wong, J. (2000) 'Repetition in conversation: a look at "first and second sayings"', *Research on Language and Social Interaction*, 33 (4): 407–424.
- Wray, A. (2002) *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.