International Journal of English Studies

IJES

UNIVERSITY OF MURCIA

www.um.es/engphil/ijes

And she's like it's terrible, like: Spoken Discourse, Grammar and Corpus Analysis

SVENJA ADOLPHS AND RONALD CARTER*
University of Nottingham

"Perhaps the greatest single event in the history of linguistics was the invention of the tape recorher, which for the first time has captured natural conversation anh made it accessible to systematic study".

Michael Halliday (1994) «Introduction» to *An Introduction to Functional Grammar* (2nd Ed) p. xxiii

ABSTRACT

This paper argues for the imporiance of teaching frequent words in English and for using computer corpora as a guide to decisions over which words to teach. The article contains a case study of a word which is frequent in both written and spoken English but more frequent in spoken English. The use of a spoken corpus raises complex questions conceming the teaching of grammar, especially frequent words in a 'discourse grammar' and these are discussed in relation to evidence of contexts of use, the needs of the learner and the use of authentic language data in the foreign language classroom.

KEYWORDS: teaching, frequent words, computer corpora, spoken corpus. discourse grammar. contexts of use

Address for Correspondence: Svenja Adolphs and Ronald Carter; School of Ennglish Studies; The University of Nottingham; University Park; Nottingham NG7 2RD; Ronald.Carter@nottingham.ac.uk; Svenja.adolphs@nottingharn.ac.uk

O Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

I. INTRODUCTION

In this paper our aim is to explore the relationship between language and discourse, especially spoken discourse and applications of such work within the context of English Language Studies. We examine data froin a five-inillion-word computerised language corpus —the CANCODE spoken English corpus. [CANCODE stands for 'Cambridge and Nottingham Corpus of Discourse in English']. The corpus was developed at the University of Nottingham, UK between 1994 and 2001, and was funded by Cambridge University Press O, with whom sole copyright resides. The spoken data were recorded in a wide variety of mostly informal settings across the islands of Britain and Ireland and then transcribed and storcd in computer-readable forni

1.1. What is a computerised language corpus?

A computerised language corpus is a collection of texts stored in electrotiic fortnat. Information about the language in the corpus is inade accessible through software desigtied to analysc patterns of language. For exainple, coniputerised language corpora can give inforrnation about the frequency of words in the corpus, the most common partnerships formed by the words with other words, the different uses of such patterns in speech and writing and the different grammatical structures found in different varieties in the corpus such as newspaper or legal language.

Most latiguage corpora in the world are assembled with the aim of making statements about language which can be statistically supported. Examples in English are the 400m word Bank of English, held at the University of Birmingham, UK and the 100m word British National Corpus (BNC). These and other corpora have proved invaluablc in the construction of authentic reference materials such as dictionaries for learners of English. Both these corpora contain spoken samples but contaiti maitily written data and there is still a tcndency for written language to predominate in computerised corpora because such data are so much easier to collect.

In spite of trends to ever larger, multi-million-word corpora and associated quantitative analysis, in the case of CANCODE the main aim has been to construct a corpus which can allow both quantitative and qualitative investigation. The data have been carefully collected and sociolinguistically-profiled with reference to a range of different speech genres and with an emphasis on everyday communication. The corpus has bcen designed with a particular aim of relating grammatical and lexical choice to variation in social contcxt and is also used in connection with a range of teaching projects, being especially concerned with differences between spoken and written language (Bex 1996, Carter and McCarthy. 1997; Carter, Hughes and McCarthy, 2000). What all these corpora havc in common is a concern with language as it is really used. They reinforce a tradition of cxamining how language is authentically and actually used rather than 'armchair'conceptions of language use in which a linguist tests hypotheses based on made-up or invented examplcs.

O Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

II. TYPES OF SPEECH IN THE CORPUS

The data collected for the CANCODE project were classified along two main axes according to **CONTEXT TYPE** and **INTERACTION TYPE**. The axes were selected with the aim of providing frameworks which are neither too broad nor too narrow. The classification scheme emerged both pre- and post-hoc in that the researchers had presuppositions concerning the contexts in which they wanted to have evidcnce of language use and yet had to develop the categories in response to the enierging data bank. There were no prior conceptions concerning instances of creative language use since that was not a primary concern of the project in its earliest stages.

11.1. Context type

This axis of categorization reflects the interpersonal relationships that hold between speakers. Four broad types were identified : *intimate,* socialising professional and transactional. (A fifth somewhat narrower category embracingpedagogic contexts to support the teaching and learning underpinning of CANCODE is not considered here). The categories embrace both dyadic and multi-party conversations. In multi-party conversations in particular it was initially thought to be problematic that relationships, especially changing relationships or relationships affected by new members joining the group, might be difficult to monitor, but a strong tendency has existed for speakers to converge towards one interaction type in their linguistic behaviour. For exainple, two intimates sharing a common place of work will adopt a 'professional' attitude in the company of colleagues. To safeguard against possible misinterpretation by the analyst, information on speaker relationships is provided in the majority of cases by the person contributing the data to the corpus. An assessment of speakers' own goals thus remains central to the analysis.

An *intimate* relationship is a private rclationship which typically (but not exclusively) centres round cohabitation and wherc speakers can be assumed to be linguistically niost 'off-guard'. All participants in a conversation must belong to the intimate sphere for the text to be categorized thus. So, for example, a conversation between two or more intimates and the family doctor on a home visit will not be 'intiinate' but 'transactional'.

The *professional* category refers to the relationship that holds between people who are interacting as part of their regular daily work. Tlie spcakers in a professional encounter need not be peers but they do nced to share either a profession or a regular place of work. So-called 'casual' talk at work is also included in this category, based on the assumption that colleagues retain the same professional interpersonal relationships whether they are discussing work matters or not. Of course, it is rccognized that colleagues can also be friends in which case their conversations could be classed as 'socialising'.

An important charactcristic oftlie transactional category is that often there is no previous relationship established between speakers. If the 'intimate' relationship is the niost private, the 'transactional' is the most public —which is one of the reasons why transactional data is

O Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

relatively easy to acquire. The transactional category includesjob interviews, asking apasser-by for information, goods and service encounters and so on.

Typical contexts for *socialising* are recreational settings such as sports clubs and pubs, as well as political, environmental, religious and other group meetings. Note, however, that it is the relationship between speakers, that is, their wish to communicate at this level, which qualifies data for inclusion in the category, and not the particular environment in which the recording is made. So, for example, a married couple engaged in private conversation in a pub will remain 'intimate'. Two couples in a similar setting, however, are more likely to conform to a 'socialising' text.

Although there are points of overlap between categories, the relationship categories do represent, albeit roughly, a cline of 'private' to 'public' speech, with the intimate and transactional categories respectively at each end of the cline. The 'professional' category is more public than the 'socialising' category, which in tum is more public than intimate.

Along the axis of INTERACTION TYPE distinctions were made between data that are predominantly collaborative and those that are non-collaborative and, further, for the collaborative type, those which are task-oriented and those which are not.

Non-collaborative texts are those in which one speaker dominates significantly, supported by back-channelling from the other speaker(s). Typically, the dominant speaker in these texts is relating an event, telling a joke, giving instructions or explanations or professional presentations. On one level, of course, these exchanges are also collaborative, but there is a level at which they resemble narration or the unilinear, asymmetrical transfer of information, rather than dialogue. The blanket term adopted to account for such an interaction type is **information provision.**

The two other interaction types classified represent more collaborative, interactive and symmetrical speech encounters. **Collaborative idea** involves the interactive sharing of thoughts, opinions, and attitudes, while the category of **collaborative task,** as the term implies, is reserved for task-oriented communication.

Overall, INTERACTION TYPE texts have proved more difficult to categorize because of the embedding of one context-type within another. Category membership is thus allocated according to the activity that is dominant in each conversation. **A** significantly more detailed account of the CANCODE corpus and its design may be found in McCarthy (1998) where the dangers inherent in reifying the categories are also fully acknowledged.

Combining the two axes of categorization provides a matrix of twelve text types as can be seen in Figure 1, which also suggests some situations in which the text types might be found.

O Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

Figure 1: CANCODE text types and typical situations in which they might be found

	Contert-type	Interaction-type	
	Information-provision	Collaborative idea	Collaborative task
Transactional	commentary by rnuseurn guide	chatting with local shopkeeper	choosing and buying a CD
Professional	oral report at group meeting	planning nieeting at place of work	colleagues window-dressing
Socialising	telling friends about a recent holiday	rerniniscing with friends	flatrnates cooking together
lntimate	partner relating the plot of a novel	siblings discussing their childhood	couple planting a small tree in their garden

III. DEVELOPING RESEARCH AND CLASSROOM APPLICATIONS: THE EXAMPLE OF LIKE

There are many applications of this research. One of the main applications to English Language Studies of this kind of computerised corpus is to help us to identify features of spoken grammar in English which have not been previously identified in any systematic way because the evidence used for most descriptions has been written English.

Here is an extract from the CANCODE corpus with the context for the talk exchange indicated. The extract is one of several used in order to explore the provenance, distribution and function of the word *like* in spoken English. In particular the aim is to provide a description of the functions of *like* for a forthcoming grammar of English. (Carter and McCarthy, forthcoming) The grammar integrates examples from both written and spoken sources and parallel corpora are compared in order to describe differences and distinctions between spoken and written contexts. In this instance the word and its grammatical properties and functions are of particular interest because it is over five times more frequent in spoken English than in written English. It should be remembered too that most grammars of English illustrate particular grammatical forms by means of sentences and with only minimal reference to a range of different speech genres and different types of social interaction. Often, however, and this is especially the case in spoken contexts, a stretch of dialogue is needed in order fully to illustrate the meaning of items across speaking terms. In the case of *like. like* emerges as a kind of discourse marker organising the patterning of discourse and marking the nature of the interaction between the speakers.

III.1. LIKE as discourse marker

III.1.1 Reported Speech

One of the more frequent uses of *like* in spoken English is to mark direct speech. This is a relatively recent phenomenon but it is extensive, the corpus reveals, in the speech of younger, (usually under 30 years of age) speakers. *Like* stands in the place of *'said that* plus quoted

O Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

speech'. As such it often introduces speech reports. In his study of CANCODE data McCarthy (1998:161) finds that '[...] in the narrative texts in the CANCODE corpus, speech reports are overwhelmingly direct speech, and with reporting verbs in past simple (said. told) or historical present says.' One of the reasons for this is to add to the 'vividness' and 'real-time staging' (ibid) of the discourse. Furthermore, replicating direct speech adds to the authenticity of a narrative. The extracts below, all drawn from the CANCODE corpus illustrate this. They both involve the recount of anarrative. Extract 1 is drawn from aconversation between three female friends while the speakers in extract two are a couple in their twenties. (Strictly speaking, the interaction-type is information provision, although it can be argued that narratives regularly do more than provide information and narrative itself, as a speech genre, regularly gets embedded into other contexts and genres of speech).

The first extract shows the speech reporting function of the word ***like.*** A group of three women in their twenties are discussing previous events. The conversation centres around an inflatable chair.

Extract 1

<S02> I was having this hideous party last weekend and there was o blow up choir so I sat in it for o bit. I was feeling really antisocial and just really wanted to go home. And Jane ond Benny had made me come cos it's this Denise and oh er she's a hairdresser and shr hod a lot of hairdressery friends. All dressed really smartly and standing round not saying anything.
<S01> Jane is?
<S02> No Jane's friend Denise.
<S01> Oh right.
<S02> So Jane made me come because she she she'd agreed to go and so she was ***like*** *"I don't want to go there and thrrr arr all these hairdressers and me and Benny."*
[laughter]
<S02> And that was that. It was really shit and I wish I hadn't agreed to do ir. Sat in rhe chair. After five minutes I was like "Yeah. Party" and singing. Making suggestions [laughter]
<S02> Suddenly became the life and soul after sitting there.
<S01> Barmy.
[laughter]
<S01> An anti antisocial choir. Maybe I'll get one.
[laughs]
<S01> They're just so ugly.
<S03> They are hideous.

The word ***like*** is used here to report the speech of other people, as well as that of the speakers themselves. The goal of the conversation is to entertain the other speakers and to keep the conversation flowing. Other elements that add to this goal and to the vividness of the conversation are the use of strong evaluative statements ('It was really shit', 'they are hideous') and the embedding of creativity in the narrative ('An anti anti social chair').

In the next extract the speakers are talking about 'Robin', an acquaintance who is in the

O Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

habit of wearing his earphones when speaking to other people.

Extract 2

<S02> But I've found erm I I got this tiny little radio that I strap on to my my my collar and then it's got earphones. You know it S u tiny little thing. And so I have that on all day.
So from sort of six seven in rhe morning I'm listening to the rudio until five six in rhe evening. And the day seems to go u lot better.
<S01> Mm.
<S02> Er the driver must think I'm absolutely insane cos half the time I I'll be walking and I'll suddenly just burst out laughing or+
<S01> [laughs]
<S02> you can see me chuckling away. I mean I've got to the point where I really look forward to
<S01> Pity you'll end up like Robin. God he comes round and erm he comes in and he he's got his ear plugs in cos he's been cycling and he stands on the doorstep going you know that really sort of intense kind of oh
<S02> Mm.
<S01> And you say "Alright Robin how ya doing?' And he's like "Oh right" And you're sort of talking to him You just think "Take your sodding ear plugs out. 2 n d he comes in the house with them. He S still got his ear plugs in.
<S02> Really?
<S01> And he's sort of talking to you and you think "My God man you can hardly converse. You 're totally unaware of people as it is. 2 n d he's got his ear plugs in.
<S02> Yeah.

The story itself has what Eggins and Slade (1997:237) following Plum (1988) call the character of an 'anecdote'which involves 'the retelling of events with a prosody of evaluation running throughout to make the story worth telling'. The use of the word ***like*** in this extract altemates with other ways of speech and thought reporting ('and you say...', 'and he's like....', 'and you think...'). Again the rather informal use of the word ***like*** is accompanied by other features of informal spoken discourse that we find in particular in the socio-cultural and intimate categories. These are further discussed below.

111.2. Other discourse functions of like

Here is another extract from the corpus which illustrates other functions of the word. The context is *intimate* and the interaction-type is ***collaborative idea.***

(A young couple, mid-twenties, at home. <S01> male *(27);*<S02> female *(25))*

<S01> So what did you do today? Apart from watch loads of adverts and
<S02> No what did what did you do today?
<S01> What did I do today. Erm oh. Had a good day today actually. Got loads of stuff sorted out. Finished loads of odds and ends.

© Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

<S02> Did you. Like w h u i
<S01> Like my progrumme. Finished that off.
<S02> Which programme?
<S01> The computer. He says that erm there was a load che= got u list of checks (sighs) I'll start again. There's u check list of things I should have done for this programme.
<S02> Right.
<S01> And er I didn't get it. I didn't either didn't pick one up or I didn't
<S02> You weren't ihere. (laughs)
<S01> Or I wasn't there. Yeah. So I I passed it but I missed u couple of ***like*** *... really stupid ihings off.*

There are a number of interesting features of the behaviour of the word *like* in this extract. *Like* here has a fundanientally analogising function. It functions to suggest points of comparison or exeniplification even if those comparisons and exaniples are not actually drawn upon. In such cases, as in the final line in the above extract, *like* also operates to mark a pause before a statenient. The analogising function of *like* is also nianifested in phrases such as *like what* ? which serves to prompt examples and illustration as in the fourth and fifth lines above.

One reason why the word *like* cannot be examined wholly in single sentence or utterance franieworks is that the extent to which the use of the word is overlaid by other grammatical patterns may easily escape attention. For example it is interesting to note how in this example like co-occurs with two other core features of spoken grammar: ellipsis and vague language.

Ellipsis is a grammatical feature in which, most commonly subjects or subjects and verbs are not employed because we can assume that our listeners know and/or understand what we mean. It is a marked feature of spoken English grammar. (see Wilson, 2000). For exaniple:

Didn't know that film was on tonight. ***(I)***
Sounds good to me. ***(It, That)***
Lois ofihings to tell you about the trip to Barcelona. **(*There* are)**
A: Are you going to Leeds this weekend:)
B: Yes, I must. ***(go to*** *Lerds* ***this weekend)***

Vague language (see Channell, 1994) includes words and phrases such as *thing, stuff, I mean or so, or something, or anything, or whatever, sort of, kind of.* Vague language softens expressions so that they do not appear too director unduly authoritative and assertive. When we interact with others there are times where it is necessary to give accurate and precise inforniation; in many informal contexts, however, speakers prefer to convey information which is softened in some way, although such vagueness is often wrongly taken as a sign of careless thinking or sloppy expression. A more accurate term should therefore be ***purposefully vague language***. For further discussion. see Eggins and Slade, (1997); Cameron, (2001).

<SO1> do you think it is affected by your faith, like you were saying you [<S 02> mm] huve kind of moral standards of not, likr hooliganising and things I mean do you think that's because of of your faith or do you ihink that's because well because of society or whatever?

© Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

In the case of *like* in these examples it is immediately noticeable that *like* shares the same communicative territory as these forms.

<S01> What did I do today. Erm oh. Had a good day today actually. Got loads of stuff sorted out. Finished loads of odds and ends.
<S02> Did you. **Like** *what?*
<S01> **Like** *my programme. Finished that off.*
<S02> Which programme?

Like co-occurs with ellipted forms such as *hada good day today actually*, *got loads oj stuff sorted out, jinished that* and with vague words and phrases such as *loads of*, *stuff* and, to a lesser extent, *odds and ends*. The corpus also reveals that in terms of social interaction *like* has a particular provenancc in more informal encounters of the socialising and intimate type. Corpus evidence reveals a significantly lower count of uses of *like* as a discourse marker in the more formal contexts associated with professional and transactional contexts. This leads, however, to the *applied* linguistic question of how far this kind of information can be patterned into a graminar of English, cspecially a yrammar of English directed primarily at advanced learners of English. How much information do leamers needs conceming contextual usage or are the broad categories of spokcn and written sufficient for most purposes?

The following extract from a new grammar of English is in a first draft form. But it illustrates, we hopc, something of the cxtent to which descriptions of grammar need to go to provide a detailed account of the lexico-grammar of words which have significant functions and distributions in a corpus of naturally-occurring language. It will be seen that it been decided at this stage not to provide more detailed contextual information but such levels of description are being kept under review as the grammar is further trialled with leamers of English throughout the world.

Extract from *The Cambridge Advanced Grammar of English* (forthcoming) (first draft)

1. Grammatical roles of *like*

1. *Likr is used as opreposition which means 'similar to'. As opreposition it often occurs with verbs of sensation such as look, sound, feel, toste, seem.*

That looks **like** *a winner*
It tastes **like** *un alcoholic drink*
People **like** *him should be put away in prison*

2. *Like* is used as a conjunction.

The manager has involved the staff in the decision **like** *o good manager should do.*

O Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

3. *Like* is also a cornrnon verb for the expression of preferences and desires. It is very frequently used with personal pronouns.

Do you like strawberries or not?
Would you like to go to Italy?
Actually, I rather like the idea.

4. *Likr* is used as a suffix. In such uses it normally forms a hyphenated structure

*She looked ill and ivus wearing u ghost-**like** cream cloak.*

2. Like **as discourse marker**

5. *Like* is very cornrnonly used in informal spoken English. One of its most frequent uses is as a rnarker of reported speech, especially where the report involves a personal reaction or response.

So this bloke, he ivus drunk, came up to nie and I'm like 'Go away, I don't want to dance'.
And my mum's like non-stop three or four times 'Come and tell your grandma about your holiday'

6. One of the rnost frequent uses of *like* in spoken English is to focus attention usually by giving or requesting an exarnple.

*The first thing that runs through your mind is **like** meningitis, isn't it?*

7. *Likr* can be placed at the end of a clause or sentence in order to qualify a preceding staternent. It also functions to indicate that the words chosen rnay not be appropriate or adequate.

Thrn she out of the car all of a sudden like and this bike hit her right in the back.
It was u shattering, frightening like.

8. When exarnples are asked for, a cornnion structure in English conversation is *like what*

A: What did you get up to today?
B: Not u lot. There were a few computer things going on
*A: Like **what**?*

9. In sorne cases *like* acts as a 'filler', enabling the speaker to pause to think what to say next orto rephrase sornething.

*They think that like by now we should be married and if we were married then it's ok **like** to get on with your life and do what you want.*

10. *Likr* can be placed at the end of a clause or sentence in order to qualify a preceding staternent. lt also indicates that the words chosen rnay not be appropriate.

Then she got out of the car all of u sudden like and this bike hit her right in the back It was a shattering, frightening experience like.

O Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

11. *Like* is also used in the structure *(It + verh 'to be'+ like,* a phrase which introduces an example or analogy of some kind. The structure is normally followed by a clause or *an-ing* form.

It's **like** *if you go to another country you always get muddled up with the currency in the first few days.*
Like *when I go to the doctors there's always loads of people in the surgery breathing germs all over you.*

12. *Like* is commonly used in spoken English with other vague expressions such as *stuff, sort of, something.*

When we were living there as students, we'd have parties and stuff **like** *that.*

Like [extract from The Cambridge Grammar of English, CLIP, 2004/5; for a class text see Carter, Hughes and McCarthy, Exploring Grammar in Context, CLIP, 2000)]

CONCLUSION

A grammar of English which is corpus-informed, based on both written and spoken examples and which illustrates the extent to which *like* functions across sentence boundaries and across speaking turns needs to find appropriate ways of highlighting such features for learners of English. In many rcspects the description goes beyond the conventional limits of grammar and becomes an exemplification of discourse grammar. A corpus-informed spoken grammar is always to some degree pushing towards the establishment of new boundaries for a 'discourse' grammar (McCarthy, 2001).

REFERENCES AND FURTHER READING

Bex. A.R. (1996). *Variety in Written English: Texts in society, societies in text.* Routledge: Loiidoii.

Caineroii, D. (2001). *Working with Spoken Discourse.* Sage: Loiidoii.

Carter, R. (1997). *Investigating English Discourse: Language, Literacy and Literature.* Routledge: Loiidoii.

Carter. R aiid McCarthy, M. (1997). *Exploring Spoken English.* Cainbridge: Caiiibridge University Press.

Carter, R. aiid McCarthy, M. (fortlicoiniiig). *The Cambridge Advanced Grammar of English.* Cainbridge: Cainbridge Uiiiversity Press.

O Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

Carter, R., Hughes, R aiid McCarthy, M. (2000). *Exploring Grammar in Context*. Caiiibridge: Cainbridge Uiiiversity Press.

Carter, R. and Nuiiaii, D. (ed)(2001). *The Cambridge Guide to Teaching English to Speakers of Other Languages* . Carnbridge: Carnbridge Uiiiversity Press.

Channell, J. (1994). *Vague Language* .Oxford: Oxford Uiiiversity Press.

Eggiiis, S. aiid Slade, D. (1997). *Analysing Conversation.* Cassell: Loiidoii.

Halliday, M.A.K. (1989). *Spoken and Written Language.* Oxford: Oxford Uiiiversity Press.

Halliday, M.A.K. (1 994). *Introduction to Functional Grammar.* Ariiold: Loiidon.

McCarthy, M. (1998). *Spoken Language and Applied Linguistics.* Cairibridge: Cainbridge Uiiiversity Press.

McCarthy, M. (2001). 'Discourse'. In R. Carter aiid D. Nuiian (eds), *The Cambridge Guide to Teaching English to Speakers of Other Languages.* Caiiibridge: Cainbridge Uiiiversity Press.

Nuiiaii, D. (1994). *An Introduction to Discourse Analysis.* Peiiguiii: Hariiioiidswortli.

Wilson, P. (2000). *Mind the Gap: Ellipsis and Stylistic Variation in Spoken and Written English.* Harlow: Longinan.

O Servicio de Publicaciones. Universidad de Murcia. All rights reserved.