ðə dʒərnəl əv læŋgwədʒ tiʧɪŋ ænd ləmɪŋ

J L T L

dɛdəketəd tu haj kwələti risərʧ ɪn əplayd lɪŋgwɪstɪks

http://www.jltl.org/jltl/

The Journal of Language Teaching and Learning, 2014–4(1), 1-14

Leave It Out! The Use of Soap Operas as Models of Spoken Discourse in the ELT Classroom

Christian Jones*, Tania Horak**

Abstract

This study analyses spoken language from a small corpus of the popular UK soap opera *EastEnders* in order to understand the extent to which the language used may be a useful model of conversational English at intermediate levels and above. Results suggest that the spoken language used in *EastEnders* has a number of similarities to unscripted conversational language in general spoken corpora. It involves extensive use of the two thousand most frequent words in the British National Corpus (BNC) spoken lists and the most frequent words and two-word chunks are comparable to general spoken corpora and a larger soap opera corpus. The findings suggest that soap operas of this type may be a useful model of spoken language as they have more similarities to unscripted, naturally occurring conversations than dialogues often found in ELT textbooks.

Keywords: Spoken language, authentic materials, soap operas, corpora

© Association of Gazi Foreign Language Teaching. All rights reserved

1. Introduction

The benefits of using authentic materials, which we can broadly define as materials which 'fulfil some social purpose in the language community' (Little & Devitt, 1989, p. 25) and are not specifically designed for use in the EFL/ESL classroom, have long been discussed within ELT. Many researchers have sought to show the advantages of these materials upon language learning and learner motivation (e.g., Gilmore 2011; Peacock, 1997), although the inherent advantages of authentic materials have also been questioned (e.g., King, 1990). There has also been a debate revolving around definitions of 'authentic'. One suggestion has been that authenticity lies in the interaction with materials and not the materials themselves (e.g., Widdowson, 1998) while others (e.g., Al-Surmi, 2012) have suggested that there is a distinction between authentic and natural materials. Al-Surmi (2012) suggests that something authentic (i.e. not designed for teaching purposes but to fulfil some social purpose in a language community) may be more or less natural, depending upon the extent to which the materials contain features of conversation evident in spoken corpora While this debate is valid, we would suggest that terms such as 'natural' carry with them an implication of value judgment which is not always helpful or illuminating. After all, one person's 'natural' conversation may be another person's unnatural conversation. Therefore, in this article we will use broad definition of the

* University of Central Lancashire, Preston, UK. Email: cjones3@uclan.ac.uk

** University of Central Lancashire , Preston, UK. Email: thorak@uclan.ac.uk

© Association of Gazi Foreign Language Teaching. All rights reserved.

term 'authentic' as described above and take 'spoken language' to mean unscripted conversations of the type found in spoken corpora and use the term 'scripted spoken language' to refer to that found in soap operas and similar TV programmes.

A main reason for the drive toward authentic materials has been dissatisfaction with textbooks and in particular with the treatment of spoken language within them. Representations of spoken language have often been found to be overly contrived (Gilmore, 2004; McCarthy & Carter, 1994) and do not give an accurate representation of many common aspects of conversations such as repetition, ellipsis, hesitation, response tokens, discourse markers and vague language (Cullen & Kuo, 2007). The impression sometimes given in such published materials is that conversations feature overly elaborate forms of language which are always problem free and that they are constructed turn by turn as opposed to being co-constructed .

One option to help alleviate this issue is to modify recordings captured for use in developing spoken corpora and create texts and exercises based upon these recordings (e.g., Carter, Hughes, & McCarthy, 2000; McCarthy, McCarten & Sandiford, 2006). Another is to use recordings of real conversations (e.g. Carter & McCarthy, 1997) which are then transcribed and analysed. However, it is surprising how few recordings of real conversations are available with transcriptions and the motivational aspects for students of listening to audio recordings of corpus data have been questioned (Cook, 1998). A final option is to use authentic materials which replicate conversations and offer a halfway point between real recordings and textbook dialogues. One type of text which has been researched fairly extensively in this regard is the soap opera (e.g., Al-Surmi, 2012; Grant & Starks, 2001).

Although scripted, soaps are based on 'everyday' topics and the conversations are at least meant to replicate conversational English. While soaps have been compared with textbooks (e.g., Grant & Starks, 2001) few studies have taken a corpus-based approach and compared them alongside general spoken corpora in order to understand the degree of similarity and difference between soap opera dialogues and naturally occurring data. Thus, a corpus of soap opera scripts (from *EastEnders*) was compiled to address this, specifically through a focus on the following research questions:

1. What percentage of the frequent words in the soap opera data are contained in the top two thousand words from the BNC?
2. Are the most common words and chunks in this data comparable to a larger corpus of soap opera English and corpora of general spoken English?
3. Which features of spoken discourse commonly found in general spoken corpora are evident and which are missing?

2. Literature review

The first argument for at least some use of authentic materials in classes in ELT (e.g., Allwright,1979; Little & Singleton, 1991; Watkins &Wilkins, 2011; Wilkins, 1976) is the suggestion that such materials are often more motivating for learners. There have been counter arguments to this, which suggest that authentic materials can be demotivating because of their cultural and linguistic 'distance' from learners (e.g., Cook, 1998). Another argument is that authenticity is not a feature of materials but, rather, how a teacher uses the material and that they are not inherently more motivating (Widdowson, 1990, 1998). A teacher might use a newspaper story in class, for example, but change the text so it becomes a matching task or includes comprehension questions. For Widdowson, this is not an authentic use because learners are not interacting with a text in the way it was intended i.e. as a newspaper story to be read. Although it seems entirely valid to suggest that not all authentic materials will work for all learners, the arguments over definitions of authenticity seem somewhat circular and, in our view, are difficult to resolve. Therefore, as mentioned in the introduction, we take a broad definition of authentic materials, as something created for a social purpose in a language community and not for the English language classroom.

Surprisingly, there has been very little empirical classroom research which has sought to prove either the benefits or drawbacks of authentic materials. The studies that do exist seem to find that authentic materials can indeed be motivating. For example, Peacock (1997) found that authentic materials increased motivation significantly compared to textbook materials in a study of Korean EFL learners at beginner level. However, the learners that were sampled did not necessarily find authentic materials to be more interesting than textbooks. This may have been affected by the level or the common sense assertion that authentic material is not inherently better than contrived material. It is easy to pick materials which students do not like but are motivated to learn from because they know they are samples of real English.

The second argument for the use of authentic materials is that textbooks have not generally offered a realistic model of spoken language. Gilmore (2004) compared seven service encounter listening dialogues in textbooks to authentic dialogues recorded using the same opening line. In general he found that the textbook dialogues excluded many of the features of the authentic dialogues, including hesitation, pausing and overlapping turns. His results suggest that the often messy nature of real conversations has often been excluded in model dialogues, in favour of presenting grammatical or functional points. In a more recent survey, Cullen and Kuo (2007) surveyed twenty four general English textbooks at a range of levels published from 2000-2006 and found that many common features of spoken grammar were given little attention. They divided aspects of spoken grammar into three categories, A, B and C. Category A included those features which need grammatical encoding such as noun phrase heads 'This food, it's nice' or past progressive to report speech 'John was saying…'. Category B included fixed lexico-grammatical units such as discourse markers (e.g., 'well', 'I mean') or vague language (e.g., 'sort of') which cannot be changed by use of grammatical means such as inflection. Category C included non-standard forms which are frequently accepted in conversational English such as 'If I was rich…' and 'There are less people around these days' but which may be labelled as incorrect in descriptive or prescriptive grammars, due to the general bias towards standard written forms**.** Their findings show that Category B features did receive some attention in textbooks but category A received almost no attention, except at advanced levels and little attention was given to Category C. This leads them to suggest that textbook are, by and large, omitting some key features of spoken language such as ellipsis and the model they present of spoken language is a partial one when compared to data from spoken corpora. This is concerning when there is evidence that authentic materials can improve spoken communicative competence. Gilmore (2011), for instance, reports on a study comparing the use of authentic materials with the use of textbook materials for Japanese learners. His results show that the students using authentic materials (in this case, excerpts from TV comedies, dramas and so on) achieved significantly better results over time on five out of eight measures of communicative competence, which was measured in a range of tests.

Despite this evidence, it is a fact that recordings of spoken English, particularly conversations, are difficult to obtain for most language teachers and are more likely to be audio rather than video recordings, simply because it is hard to video conversations without participants knowing you are doing so and thus authenticity may well be compromised. As a result, such recordings can be difficult to place in a clear context. The scripted spoken English of soap operas may therefore be a useful 'halfway house' between spoken English and textbook dialogues. Previous research into soap operas has, above all, explored speech acts and made comparisons to either naturally occurring conversations or textbooks. For example, McCarthy and Carter (1994) analysed a section of the Australian soap *Neighbours* to examine the speech act of asking for a favour; they found that the soap dialogue was much more complex, both linguistically and in terms of the discourse organisation than the simple sequences often presented in textbooks. They also suggested that the soap dialogue contained many discourse and linguistic features which we would find in unscripted conversations. Grant and Starks (2001) took a conversation analysis approach in examining how conversations are closed in EFL textbooks when compared to fifty episodes of the New Zealand soap *Shortland Street.* They found that the closings in the soap opera data were linguistically much more varied than the

textbook models and included phrases such as 'be seeing you' and 'cheers' whereas textbooks tended to feature only 'goodbye/bye' and 'see you later' (p.45). They also found that the soap dialogues were better able to follow the typical moves involved in closing conversations, as described by Schegloff and Sacks (1973), namely that participants often closed down a topic, made a pre-closing move and then closed the conversations. In textbook dialogues, these moves were often not in evidence, leading to abrupt and pragmatically inappropriate models being given. Fahey Palma (2008) examined apologies in a fifty thousand word corpus of the Irish Soap *Fair City* and the Chilean soap *Amores de Mercado* to compare how the speech act is realised in two different languages. Contrary to the notion that speech acts are universal (Brown & Levinson, 1987), her findings show that in the Irish soap an expression of regret was the most common form of apology strategy while in the Chilean soap the 'use of verbs that formulaically and directly demand forgiveness or express an apology are the preferred strategies'(Fahey Palma, 2008). This shows that linguistically apologies do in fact vary across cultures, which suggest that for EFL/ESL learners it may be worth exploring the differences between speech acts in L1 in comparison with English.

More recently, Quaglio (2009) analysed a corpus of the American sitcom *Friends* in comparison with a corpus of conversational English. His findings show that *Friends* was similar to unscripted conversations in many respects and shared many core lexico-grammatical features. The sitcom differed in that it featured fewer instances of vague language and narratives and more instances of informal and emotional language. He suggests that these differences can largely be accounted for by the expectations of the sitcom genre. Vague language, for instance, may be more prevalent in unscripted conversations because they take place in a context shared by speakers and in sitcoms, the context is contrived and the audience are not directly involved in it. Al-Surmi (2012) has developed this analysis and taken a multi-dimensional, corpus -based approach to compare the spoken language used in a corpus of the American Soap *The Young and The Restless* with the sitcom *Friends* and with the American conversation sub-corpus from Longman Spoken and Written English Corpus (Biber, Johansson, Leech, Conrad & Finegan, 1999). Al-Surmi's findings suggest that the sitcom data have more features of spoken English as found in the conversational corpus in the areas of involved vs. informational, overt expression of argumentation or persuasion and abstract vs. non-abstract information or style, while the soap opera data were closer to the corpus data on narrative vs. non-narrative discourse (p.692). This leads him to suggest the soap operas may be more useful for modelling certain types of spoken English narrative discourse when teaching features related to narrating events, while sitcoms may be more useful for features such as non-narrative descriptive discourse. Overall, the research offers an illuminating analysis but Al-Surmi acknowledges that it is not the type of work which most language teachers would have the time to undertake and suggests that ultimately research should produce a list of TV shows for teachers which seem particularly good at offering models for specific aspects of spoken language.

Many of the studies reviewed suggest that soap opera data offer a model of spoken language which is at least closer to spoken English than the model found in many textbooks. However, the research has tended to focus on specific speech acts, rather than how soap operas in general replicate the lexico-grammatical and discourse features of unscripted conversations. Al-Surmi's (2012) paper does address this issue and the results are interesting but, as the writer acknowledges, it is not the type of research which most language teachers would be able to undertake. Lastly, few of the studies mentioned make reference to the types of levels at which we might use soap opera in the classroom. It is these gaps which this paper seeks to fill. Our intention was to analyse the data in order to find out the extent to which the sample soap replicates the lexico-grammatical and discourse features of unscripted conversations. We approached this using mainly open-access corpus tools, as a model for the kind of analysis which teachers and researchers could carry out themselves to inform classroom practice. The intention is to inform teaching at intermediate levels and above because we feel it is at these levels that learners' interlanguage will have developed sufficiently to follow this kind of material. This is not to suggest that lower level learners could not use soaps but the research was undertaken with the view that the soap operas could be used at intermediate levels and above.

3. Method

3.1 Research design

This research followed a mixed-methods approach. A small-scale focussed corpus was built and compared with larger reference corpora. Quantitative analysis was undertaken in order to ascertain frequency patterns of common words and chunks. Following this the data was analysed more holistically to look for common features of spoken grammar.

3.2 Data sources

In order to answer the research questions, a mini-corpus of *EastEnders*, the popular UK soap, was created. The corpus consists of two complete scripts from two thirty minute episodes, a number of memorable dialogues posted as episode 'tasters' on the programme website (BBC, 2012), dialogues from a 'memorable quotes' website (IMDb,2012) and eleven transcripts of episodes from a fan website (Oocities, 2012). In total, the corpus consisted of 58,142 words. The quotes used consisted of a minimum of a two part exchange and no single lines were used in order to allow analysis of scripts attempting to replicate dialogic interaction. The scripts were from two episodes in 2006 and 2007, the transcripts from the early to mid-nineties and the memorable quotes and 'tasters' from early episodes to the present day. All stage instructions were removed for the purposes of the analysis and in the case of the transcripts, spellings were standardised to ease analysis, as there was some variation and attempts to transcribe according to speakers' accents. This means that words transcribed as, for example, ' leavin'' were modified to 'leaving' and 'cup 'o tea' to 'cup of tea'.

3.3 Data analysis

The data were first analysed in Compleat Lexical Tutor(LexTutor) (2012) to discover the most frequent words and chunks and the percentage of the common words which matched the most common two thousand words in the British National Corpus (BNC, 2012). Frequency comparisons were made with three reference corpora: the Cambridge and Nottingham Corpus of Discourse in English (CANCODE) (as described in O'Keeffe, McCarthy and Carter, 2007, chapters two, three and seven), the spoken section of the British National corpus (BNC) and the American Soaps corpus (Davies, 2012). Following this, a keyword analysis was undertaken, to uncover the words which occurred with significantly greater frequency in the *EastEnders* data than the spoken section of the BNC. Finally, the data were examined quantitatively and qualitatively to explore which common features of spoken English seemed to occur frequently in the data and those which did not. This final analysis adapted the framework (of A, B and C types of features) used by Cullen and Kuo (2007) as described in the literature review. The data were examined for evidence of features of Cullen and Kuo's Category A. As mentioned previously, this is composed of those features which need grammatical encoding such as noun phrase heads or past progressive to report speech. The features chosen for our analysis were ellipsis and past progressive used to report speech. Category B included fixed lexico-grammatical units or vague language which cannot be changed by use of grammatical mean such as inflection. The features chosen here were discourse markers and non-minimal response tokens. Our Category C differed from Cullen and Kuo's because we attempted to look for typical features of conversation at the level of discourse. The features we examined here were repetition and overlapping. Each feature chosen for our analysis was felt to be a prototypical feature of conversational English and space limitations meant it would be impossible to analyse all aspects of spoken language which Cullen and Kuo mention.

4. Results and Discussion

RQ1: What percentage of the frequent words in the soap opera data are contained in the top two thousand words from the BNC?

Table 1. Soap opera data and the first two thousand (K2) words from the BNC

Freq. level	Families	Types	Tokens	Coverage (tokens) %	Cum %
K1 words	838	1650	58,139	92.18	92.18
K2 words	512	695	1,704	2.70	94.88
K3 words	305	350	621	0.98	95.86

This shows that almost 95% of the words used in the first two thousand most frequent words in the BNC, which we would expect the majority of learners at intermediate levels to have a firm understanding of, are found in the soap opera data. This does not quite reach the figure of 95 % coverage which is often said to be required for comprehension of reading texts (Hu & Nation, 2000) but it is still clear that the majority of the words in the corpus come from the first thousand in the BNC and as such it can be judged as a reasonable and attainable model for intermediate learners. Three examples of words from the corpus are shown in table two below.

Table 2. Examples of words found at first 3 K levels

K1 words	A, about, act
K2 words	Background, banged, bathroom
K3 words	Canal, cans, casual

RQ2. Are the most common words and chunks comparable to a larger corpus of soap opera English and a corpus of general spoken English?

Table 3 below shows the most frequent twenty five words in the *EastEnders* corpus, in comparison with the BNC spoken corpus (10 m words), the soap opera corpus (10 m words) and CANCODE (5 m words)

Table 3. The twenty five most frequent words in four corpora

RANK	*EastEnders* data	RANK	BNC spoken corpus	RANK	CANCODE	RANK	US SOAPS
1.	YOU	1.	THE	1.	THE	1.	YOU
2.	I	2.	I	2.	I	2.	I
3.	TO	3.	YOU	3.	AND	3.	TO
4.	A	4.	AND	4.	YOU	4.	THE
5.	THE	5.	A	5.	IT	5.	THAT
6.	IT	6.	'S	6.	TO	6.	IT
7.	AND	7.	TO	7.	A	7.	AND
8.	WHAT	8.	OF	8.	YEAH	8.	N'T
9.	YEAH	9.	THAT	9.	THAT	9.	A
10.	ALL	10.	N'T	10.	OF	10.	DO

11.	ME	11.	IN	11.	IN	11.	WHAT
12.	WELL	12.	WE	12.	WAS	12.	OF
13.	THAT	13.	IS	13.	IT'S	13.	ME
14.	OH	14.	DO	14.	KNOW	14.	IS
15.	OF	15.	THEY	15.	IS	15.	KNOW
16.	KNOW	16.	ER	16.	MM	16.	THIS
17.	BE	17.	WAS	17.	ER	17.	HAVE
18.	RIGHT	18.	YEAH	18.	BUT	18.	HE
19.	NO	19.	HAVE	19.	SO	19.	WE
20.	FOR	20.	WHAT	20.	THEY	20.	FOR
21.	DO	21.	HE	21.	ON	21.	IN
22.	DON'T	22.	THAT	22.	HAVE	22.	JUST
23.	IN	23.	TO	23.	WE	23.	MY
24.	JUST	24.	BUT	24.	OH	24.	NOT
25.	ON	25.	FOR	25.	NO	25.	WAS

Although the frequencies vary between the corpora, there are clearly similarities. The most common words contain few items which contain propositional meaning and many items which act as function words such as 'to', 'of' and 'me'. The high frequency of 'I' and 'you' as opposed to 'he' and 'she' shows that the *EastEnders* dialogues are similar to general conversations in that they concern the speaker and the person they are addressing most frequently. What is interesting is the absence of response tokens such as 'Mm', and hesitation devises such as 'Er' in the *EastEnders* , BNC or larger soap opera corpus while both occur with high frequency in CANCODE. This may reflect, to a degree, the scripted nature of the dialogues. Characters do not need to react to the ongoing discourse as they know the line which is coming next and are waiting for their cue. In the BNC, the absence of such markers is likely to reflect the fact that is it made up, in part, of prepared spoken language in the form of public lectures and so on. The keyword analysis conducted compares the *EastEnders* data to the spoken section of the BNC, to uncover which words occur with significantly higher frequency in *EastEnders.* This produces a keyness factor, a calculation which demonstrates how much more frequent a word is in one data set when compared with a general reference corpus. The higher the figure, the more 'key' it is in the data set under investigation. Lextutor produces a long list of keywords but for the purpose of this article, only those with a keyness factor of 50 or more (see Table 4) were analysed, as Chung and Nation (2004) recommend this is an effective cut off point.

Table 4. Keywords with a keyness factor of 50+

(1)	1155.00	halo
(2)	825.00	mistletoe
(3)	660.00	jack
(4)	495.00	derrick
(5)	495.00	chippy
(6)	495.00	gaff
(7)	495.00	weight
(8)	179.95	valentine
(9)	165.00	scrubber
(10)	101.54	blossom
(11)	99.00	pamper
(12)	75.58	wick
(13)	61.88	thug
(14)	55.00	uptight

While some of these words , such as 'scrubber', 'chippy' and 'thug 'reflect the relative informality of *EastEnders*, others such as 'valentine' and 'halo' reflect the topic of some episodes, based around Valentine's Day, Easter and Christmas. Others such as 'Derrick' reflect the fact that characters use each other's names a great deal and there are a group of names which are also homographs of certain nouns and verbs. As this list is relatively short, the data suggest that *EastEnders* does not contain a large amount of lexis which learners at intermediate levels will struggle with and the keywords that do exist could easily be glossed or pre-taught. Table Five shows the most common two-word chunks in the data when compared with CANCODE

Table 5. Most common two word chunks (number of occurrences in brackets)

EastEnders	CANCODE
001.**[158]** I DON'T	001.**[28,013]** YOU KNOW
002.**[155]** YOU KNOW	002.**[17,158]** I MEAN
003.**[144]** DO YOU	003.**[14, 086]** I THINK
004.**[106]** IN THE	004.**[13,887]** IN THE
005.**[106]** ALL RIGHT	005.**[12,608]** IT WAS
006.**[100]** I MEAN	006.**[11,975]** I DON'T
007. **[95]** A BIT	007.**[11,048]** OF THE
008. **[95]** I WAS	008. **[9,772]** AND I
009. **[94]** TO BE	009. **[9,586]** SORT OF
010. **[88]** IF YOU	010. **[9,164]** DO YOU
011. **[87]** WELL I	011. **[8,174]** I WAS
012. **[79]** I THINK	012. **[8,136]** ON THE
013. **[78]** WANT TO	013. **[7,773]** AND THEN
014. **[77]** GOT TO	014. **[7,165]** TO BE
015. **[77]** I THOUGHT	015. **[6,709]** IF YOU
016. **[74]** I KNOW	016. **[6,614]** DON'T KNOW
017. **[70]** IT WAS	017. **[6,157]** TO THE
018. **[67]** YOU WANT	018. **[6,029]** AT THE
019. **[66]** HAVE A	019. **[5,914]** HAVE TO
020. **[65]** TO SEE	020. **[5,828]** YOU CAN

It is clear that many of the most common chunks in the *EastEnders* data have some similarities to the CANCODE data. 'You know', 'I mean' and 'I think' are frequent in both corpora, for example. What is also striking is the higher frequency of 'I know' in the *EastEnders* corpus and it worth exploring why this occurs with higher frequency and also comparing it with those chunks which occur with similar frequency. In the *EastEnders* data the chunk 'I know' seems to be highly frequent because it is used to mark what a character is saying and signal that they are being 'genuine' or that they understand or have an acceptance of something:

Extract 1*: I know*

It can't kill love. And I got that, Jim**. I KNOW** I'm loved. You can't tell me what to
you learn from your mistakes and move on. One thing **I KNOW** is you don't go out for hamburger wh
I'm not drunk. I know what I'm saying. I love you. What? **I KNOW** it now, Stacey. I always have..
see you in the Vic later? I'll get her there, don't worry. **I KNOW** it's the thought that counts but.
He's been giving me the silent treatment. I'm sorry. **I KNOW** I've made things difficult between

In the CANCODE corpus, 'I know' is not as frequent as 'you know' where it is often used as a discourse marker to indicate shared knowledge or as a pause marker (O'Keeffe et al., 2007, p.71). It is not used in this way in *EastEnders* as often, perhaps largely due to the contrived nature of the interaction. Instead it is commonly used to introduce a 'pearl of wisdom' or as a rhetorical question, to mark the fact that the character is going to say something important.

Extract 2: *You know*
and we're gonna have a little chat. Nice one. From Abi. **YOU KNOW** what women are like about Valentine's
Take me to Ian's. I want to see if I can help. I know. **YOU KNOW** what? I know you were in prison. I'm
always have a habit of coming back and haunting you. **YOU KNOW** what? I probably could manage a bit o
still stuck in your local with your ex, aren't you? And **YOU KNOW** what? Maybe that's where you should
we're lucky tonight we've no women in tow. And **YOU KNOW** why? Cos it never works out, son

'I mean', on the other hand, seems to function largely as a discourse marker in the *EastEnders* corpus, just as it does in CANCODE. It is largely used to mark the fact that a character wishes to reformulate or clarify something they have just said:

Extract 3: *I mean*
Is that, is that all I am to you, Ian? **I MEAN** is that all I mean to you
Sorry? The lack of consideration **I MEAN** it doesn't take much
And I'm the one supposed to tell her **I MEAN** it's not fair though, is it, Nat.

RQ3. Which features of spoken discourse commonly found in general spoken corpora are evident and which are missing?

Category A features (Ellipsis, 'X was saying')

Ellipsis is very common in the data and occurs in many of the dialogues between the characters. Partly it seems to be used to mark informality and signal friendship and familiarity but also because it fits many of the situations. For example, being over elaborate would not be required in many of the situations featured in *EastEnders* such as buying things from the local shops or café. In this sense, the dialogues are similar to natural recordings, where it has been shown that situational ellipsis is prevalent (Carter & McCarthy, 1997, 2006). The three short dialogues below demonstrate this, although we found many more in the data.

Extract 4: *Examples of ellipsis*
S1: Calmed down yet?
S2: Oh yeah. Look at me. Total calmness.
S3: I don't know. There's just something different about you.
S4: Like what?
S3: A glow maybe
S5: A sponge, some chocolate chip cookies and something with cream in it.
S6: No fairy cakes?
S5: Just stick 'em in a bag!

The use of past progressive to report speech, however, was almost totally unused in the data with only the following example found:

Extract *5: Example of past progressive*
S1: Dot was just saying it's her anniversary today.
S2: Congratulations.
S3: That's more than I got from Jim

Instead, a search for the word 'saying' revealed that it tended to be used in present progressive form (sometimes displaying ellipsis) to either offer explanation of what a character expresses or to check what another character utters:

Extract 6: *Saying*
People like to see a friendly face behind the bar. **SAYING** mine aint? I think that hair lacquer's
You can have it if you like. What's this? It's me **SAYING** you aint bad. For
I'm not drunk. I know what I'm **SAYING**. I love you. What? I know it now

When characters report speech they tend to use 'said'

Extract 7: *Said*
Alright, Ian, take your time. Peter **SAID** he was in bed last night.
A friend of mine has a cottage in Suffolk. She **SAID** I can use it any weekend
Really, do we? Something Den **SAID**, actually. That everyone has a skeleton

This suggests the *EastEnders* dialogues do not mirror this common feature of spoken language.

Category B (discourse markers, response tokens)

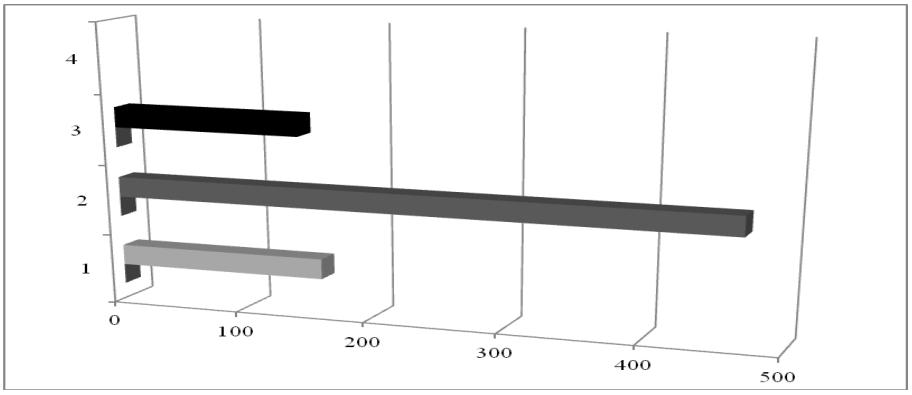
The data was examined to see if two common discourse markers 'Oh' and 'well' were used in a similar way to a general reference corpus. These items were chosen because they occur with high frequency in most corpora of spoken English (CANCODE, for example lists 'Oh' as the 24th most common word, 'well' as the 27th most common, O'Keeffe *et al*., 2007, p.35/65). As the frequency counts in Table 1 (above) show, each discourse marker (DM) did occur often in the data with 'well' being the most frequent, followed by 'oh'.

The following examples show each DM being used in context:

Extract 8: *Oh* and *Well*
Couple of 'loving cups'? Just a beer,please. **OH** come on, I'm pushing the boat out
you can manage to get rid of Bert and Jay for the evening **OH** come on, Mum
I.said we'd look after the girls tonight. You what? **OH** great. Roast chicken!
grateful for your feedback. What do you want my boots for? **WELL** I ain't using these, finest hand
did that alright. It was just meant to be a bit of fun but **WELL** I wouldn't want you to go
Love you too I don't think I'm immature **WELL** I' m telling you you are.

We can see that these examples, together with use of common DMs such as 'I mean' suggest that we can say that the *EastEnders* data are similar to spoken language found in CANCODE to a reasonable degree. When we examine the use of response tokens, a slightly different picture emerges. We have already noted that minimal response tokens such as 'mm' are quite rare in the *EastEnders* data. According to O'Keeffe *et al.,* (2007) the most common non-minimal response tokens in British and American English are 'good', 'right' and 'really'. All three occur in the *EastEnders* data with the frequencies shown in Figure 1 below.

Figure 1. Use of three response tokens in the *EastEnders* corpus (good = black, right= dark grey, really= light grey)



However, it is notable that it is only 'really' which is consistently used as a response, as in the following examples:

Extract 9: *Really*
okey you in at that new treatment place up the high street. **REALLY**? After yesterday... You What about you?
No, I'm afraid you've got me beat there. **REALLY**? Maybe we should change
You remind me of two girls I used to know. **REALLY**? What mates of yours? Yeah...

'Right' occurs with a high frequency but is almost never used as a response. Instead, it forms the common chunk with 'all right' or is used with a propositional meaning to suggest that something is correct or as part of a prepositional phrase.

Extract 10: *Right*
Reckon that'll do me tonight. You were **RIGHT** about the Vic probably be full of
My nan's a battleaxe. My cousin Mo's all **RIGHT** but my cousin Zoe, wait till you get a
He was right there **RIGHT** by those bushes. It's his. No...

'Good' is used in some instances as a response and in some cases with an adjectival meaning. The sample below shows each of these uses:

Extract 11: *Good*
How're things at home. Okay? Yeah. They're fine. Really. **GOOD**! You and Nigel getting on a bit better
We do a good enough job, and we get the permanent one easy. **GOOD**, I'm glad. Oh, me too.
All right.. I took your advice. Went out for a walk. Oh. **GOOD**. And? Now I'm back.
Not like this place. Hm hm! Still, you got to work with the **GOOD** people if you want to improve.
Especially when you've got a perfectly **GOOD** place of your own. What place?
I think it's a good present! Well, it isn't. Huh? No. A **GOOD** present Ian, it, is something that's special

The reasons for the limited use of 'good' and 'right' may again be because characters do not need to respond simply because they know what is coming next. 'Really' may differ because its use can signal that something dramatic or interesting has been said, rather than the more mundane use of 'right' to signal that simply one character is following the other. In spoken English, the absence of tokens such as 'right' can also signal a lack of interpersonal awareness (i.e. the listener is not actually listening) but in soap, this type of interpersonal engagement is clearly not as important as in unscripted exchanges. While scriptwriters clearly aim at dialogue mimicking real life, realistic interaction would not be entirely conducive to maintaining pace and clarity for viewers.

Category C. Overlapping and repetition

There was very little evidence of either of these features in the data. Largely, this would seem to be the result of the fact that *EastEnders* is a scripted drama and actors know what is coming. Therefore, the kind of overlapping which is a common part of English discourse is not really in

evidence, again probably to maintain clarity in the dialogues. Equally, the scriptwriters are perhaps unaware of this common feature of real speech. This finding is similar to Quaglio's (2009) analysis of the sitcom *Friends.* He suggests that the restrictions of the genre may override the need to exactly mimic unscripted conversations, which often features latched and co-constructed turns (Carter & McCarthy, 2006). In *EastEnders,* each episode is only thirty minutes long and there is a clear need for characters to say their lines, move the plot along and keep the audience interested. Should turns overlap a great deal, this may be harder to achieve in the time allowed for each episode. It is also the case, as Quaglio (2009) notes in regard to *Friends*, that the conversations in *EastEnders* are based on how the scriptwriter perceives spoken English and are unlikely to be based on analysis of spoken corpora.

Tannen (1987) suggests that repetition is pervasive in conversation within and across turns. The reason for this is that in general it aids coherence and cohesion and allows speakers to produce language more effectively, listeners to comprehend language more easily and for speakers to interact more effectively on an interpersonal level. This is despite the fact that many non-linguists view repetition negatively, 'as any use of language that does not convey information is seen as superfluous and therefore bad' (Tannen, 1987, p.585/6).This may also be as a result of applying the norms of some genres of written language (e.g. academic writing), where repetition can be viewed negatively, to spoken language.

Repetition does occur in the *EastEnders* data but not with the same frequency as it might occur in unscripted conversation. The following sample shows some evidence of repetition:

Extract 12: *Example of repetition*
S1: *Oh, no! You're here! You haven't answered any of my texts!*
S2: Well, I wasn't sure which one to reply to. There were fourteen of them. Well, fifteen now.
S1:The one about dinner.
S2:Oh, yeah.
S1:I was wondering... if... you might like... a home-cooked meal sometime.
S2:Oh, that'd be lovely. What we having?
S1: Sausage surprise! I'm known for it around here.

Although there are some examples of repetition here, in general it would seem that there are fewer instances in *EastEnders* because its scripted nature means it is not required as an aid for production or comprehension. The most notable absence, which we can see in this example, is that speakers do not tend to repeat what others have said. Tannen (1987) suggests that this function of repetition is largely interpersonal. Speakers may repeat what others have said to, for instance, show they are listening or are interested in what has been said to them. This is similar to the relative absence of response tokens, as noted above. In a soap opera, characters do not show they are listening, have understood or are interested as often as in unscripted conversation because they know what is coming next.

5. Conclusion

This study has shown that soap opera dialogues share some important characteristics of conversation including many of the most frequent words from the BNC, ellipsis, discourse marking and common chunks from CANCODE.

Based on this evidence, we can suggest that soap operas can act as a bridge between, on the one hand, often unnatural textbook dialogues and, on the other, recordings of unscripted conversations, the latter of which may be inaccessible to teachers or difficult to comprehend for learners with a developing interlanguage. The dialogues used in this particular soap opera will need supplementing to give a clearer model of features of conversational English, including the use of response tokens such as 'mm' and 'right' , the use of 'X was saying' to report speech and the tendency for spoken

English to feature a great deal of repetition, overlapping and co-constructed turns. However, it is clear that the *EastEnders* data do offer some of the common features of conversation and as such could be used as a useful model of conversational language in classes. The *EastEnders* dialogues could easily be used as listening comprehension or to contextualise and raise awareness of features such as ellipsis**.**

Naturally, there are several limitations to this research. The soap opera in this article is British and may not be appropriate for all ELT contexts. To address this, the same type of analysis could be undertaken with another soap which a researcher or teacher feels is most appropriate to a particular cultural context. This would certainly include English-medium soap operas where English is being used as a lingua franca, as these may equally contain a useful model of successful conversational English. It would also be helpful to trial the use of soap opera materials in a classroom research project which could assess the effectiveness of a soap opera in comparison to textbook materials as a means of developing spoken communicative competence. Gilmore's framework (2011) could easily be used as a template for this kind of research and the results could help a range of teachers to evaluate the use of soap operas as a model of everyday spoken language for their students.

Biodata

Christian Jones is a Senior Lecturer in TESOL at the University of Central Lancashire and has previously worked as a teacher and teacher trainer in Japan and Thailand. His main research interests are in spoken discourse analysis, corpus-informed language teaching, lexis and the pedagogical treatment of spoken grammar.

Tania Horák is a Lecturer at the University of Central Lancashire, co-ordinating EFL programmes and teaching on TESOL teacher training programmes. She has previously worked in the Czech Republic, Bangladesh, Lithuania, Hong Kong and Germany. Her research interests lie in foreign language testing and assessment, academic writing, and corpus linguistics.

References

Allwright, R. (1979). Language learning through communication practice. In C. Brumfit, & K. Johnson, *The Communicative Approach to Language Teaching* (pp. 167-182). Oxford: OxfordUniversity Press.

Al-Surmi, M. (2012). Authenticity and TV shows: a mutlidimensional analysis perspective. *TESOL Quarterly 46(4)*, 671-694.

BBC. (2012, November 1). *Script Peeks* . Retrieved from http://www.bbc.co.uk/programmes/b006m86d/features /script-peeks

BBC. (2012, November 1). *Writers Room.* Retrieved from http://www.bbc.co.uk/writersroom/scripts/

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English.* London: Longman.

Brown, P., & Levinson, S. (1987). *Politeness: Some Universal in Language Usage.* . Cambridge: Cambridge University Press.

Brtish National Corpus (BNC) (2012, October 3). Retrieved from http://www.natcorp.ox.ac.uk/

Carter, R., Hughes, R.., & McCarthy, M., (2000). *Exploring Grammar in Context.* Cambridge: Cambridge University Press.

Carter, R., & McCarthy, M. (1997). *Exploring Spoken English.* Cambridge: Cambridge University Press.

Carter, R., & McCarthy, M. (2006). *Cambridge Grammar of English.* Cambridge: Cambridge University Press.

Chung, M., & Nation, P. (2004). Identifying technical vocabulary. *System 32(2)*, 251-263.

Compleat Lexical Tutor. (LexTutor) (2012, October 2). Retrieved from http://www.lextutor.ca/

Cook, G. (1998). The uses of reality: a reply to Ronald Carter. *ELT Journal 52(2)*, 57-63.

Cullen, R., & Kuo, I. (2007). Spoken grammar and ELT materials:a missing link? *TESOL Quaterly 41(2)*, 361-386.

Davies, M. (2012, October 4). *Corpus of American Soap Operas.* Retrieved from http://corpus2.byu.edu/soap/

Fahey Palma, M. (2008, November 1). *Speech acts as intercultural danger zones: a corpus-based analysis of*

apologising in Irish and Chilean soap operas. Retrieved from http://www.immi.se/intercultural/nr8/palma.htm

Gilmore, A. (2004). A comparison of textbook and authentic interactions. *ELT Journal 58(4)*, 363-374.

Gilmore, A. (2011). "I Prefer Not Text": Developing Japanese Learners' Communicative Competence with Authentic Materials. *Language Learning 61(3)*, 786-819.

Grant, L., & Starks, D. (2001). Screening appropriate teaching materials:closings from textbooks and television soap operas. *IRAL 39*, 39-50.

Hu, M., & Nation, I. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language 13(1)*, 403-430.

IMDb. (2012, November 5). *Eastenders Memorable Quotes.* Retrieved from http://uk.imdb.com/title/tt0088512/quotes

King, C. (1990). A linguistic and a cultural competence: can they live happily together? *ForeignLanguage Annals 23(1)*, 65-70.

Little, D., & Devitt, S. S. (1989). *Learning Foreign Languages from Authentic Texts: Theory and Practice.* Dublin: Authentik.

Little, D., & Singleton, D. (1991). Authentic texts,pedagogical grammar and language awareness in foreign language teaching. In C. James, & P. Garret, *Language Awareness in the Classroom* (pp. 123-132). London: Longman.

McCarthy, M., & Carter, R. (1994). *Language as Discourse:Perspectives for Language Teaching.* Harlow: Longman.

McCarthy, M., McCarten, J., & Sandiford, H. (2006). *Touchstone Intermediate Sudent's Book.* Cambridge: Cambridge University Press.

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

Oocities. (2012, November 10). *EastEnders.* Retrieved from http://www.oocities.org/tvtranscripts/eastenders/index.htm

Peacock, M. (1997). The effect of authentic materials on the motivation of EFL learners. *ELT Journal 51(2)*, 144-156.

Quaglio, P. (2009). *Television dialogue: The Sitcom Friends vs Natural Conversation.* Amsterdam: John Benjamins.

Schegloff, E., & Sacks, H. (1973). Opening up closings. *Semiotica 8*, 289-327.

Tannen, D. (1987). Repetition in conversation:towards a poetics of talk. *Language 63(3)*, 574-605.

Watkins, J., & Wilkins, M. (2011). Using YouTube in the EFL classroom. *Language Education in Asia 2(1)*, 113-119.

Widdowson, H. (1990). *Aspects of Language Teaching.* Oxford : Oxfrod University Press.

Widdowson, H. (1998). Context, community and authentic language. *TESOL Quarterly 32(4)*, 705-716.

Wilkins, D. (1976). *Notional Syllabuses.* Oxford: Oxford University Press.