

KEYNOTE PRESENTATIONS

Spoken English/Written English - From Corpus to Curriculum to Classroom

Ronald Carter,
University of Nottingham



I must confess I feel a touch fraudulent standing here. I haven't done a lot of work in relation to English as an additional language. Most of the work I have done has involved descriptions of English, that I hope have particular relevance for EFL, ESOL, English as an International Language and so on. But this is a very specific context and I'm therefore a little bit wary about offering insights that may not key directly and relevantly into your concerns, so I've decided therefore to raise some general questions about the relationship between spoken English and written English, which I hope have implications for all languages.

My starting point is that the better we understand these things and the better we understand the differences and distinctions between spoken languages and written languages, the better we can teach them. My basic position is that we know a lot about written language, we know a lot less about spoken language.

We know a lot about this kind of thing. We know a lot about how different levels of formality can be constructed in English. So here, the kind of sentence

that no-one would ever really produce, except in fairly limited contexts, you can see, can't you, and we know this already in a sense, that if you want to have greater formality in written English, certainly the way in which you do it principally, is to build up noun phrases.

- *Glass cracks more quickly the harder you press on it.*
- *Cracks in glass grow faster the more pressure is put on.*
- *Glass crack growth is faster if greater stress is applied.*
- *The rate of glass crack growth depends on the magnitude of the applied stress.*
- *Glass crack growth rate is associated with applied stress magnitude.*

(Halliday, 1989)

So in the first example, "glass" is the only noun, phrase, noun, "head" there, but in the second, "cracks", which is a verb in the first sentence, becomes a noun in the second sentence, and so on and so forth. Until you get from "glass cracks more quickly the harder you press on it" through the levels of formality, you get to "glass-crack growth rate is associated with applied-stress magnitude". We're actually very good at this. I think teachers do understand very well different levels of formality within written language. We know the whole process of what is called here 'nominalisation', taking different parts of speech, turning them into nouns, building phrases around them, and that leads to greater formality, the kind of formality we often see in specific academic contexts of use.

We're also quite good at identifying structures that are specific to written language – here's one:

First staged at the Glasgow Citizens in 1994, and described by Williams as being a 'comedy of death', the play sees Everett cast brilliantly against type as the rich dying widow Flora Goforth.

I defy anyone to have ever said this. If someone asks you "Did you enjoy the play last night?", I'd suggest you're unlikely to answer: "First staged at the Glasgow Citizen's in 1994 and described by Williams as being a comedy of death, the play sees Everett cast brilliantly against type as the rich dying widow Flora Goforth." Or if you did, I'd suggest you have very few friends.

That type of structure, which involves a non-finite past participle clause, would you believe, is hardly ever heard in spoken English. So, my position is: we're good at this, we understand intuitively and explicitly and analytically, what goes into the organisation of written language and, very generally, we are good at teaching it and we are getting better all the time.

What has also helped us is the advent of what I consider to be one of the major developments in Applied Linguistics over the course of the last 20 years, the development of large, multi-million word databases and the introduction of corpus-based approaches to the description of language. And a corpus, which is not a dead body, but essentially a large collection of texts, very often electronically stored, with software available to enable you to do very quick searches across those databases, are now beginning to inform, as you know, more and more dictionaries and grammars and so on. I have a number of words of caution about corpora to advance later in the talk, but I do think they have led to what I believe is very important, and as I said at the beginning, more precise descriptions of how the language works. We simply have now much more evidence available to us to back up intuitions and to provide examples, concrete, real world examples of how language actually operates.

We now have a British National Corpus, which used to be hard to access, but now, if you type in “British national corpus” into Google, you can get information very quickly about how it operates, you can get a CD with very usable software, which enables you to work in and with 4 million words, 3 million words of written English and a million words of spoken English. The whole corpus itself, if you want to access this, is a hundred million words.

There are other corpora that are easily available. I won’t go through all of this, but if you simply type “corpus” into Google, it will give you the list. But MICAS is very interesting, this is the Michigan Corpus of Academic Spoken English and is one of the few examples of spoken corpora around in academic settings, and I do suggest you have a look at that, because it’s available on line and it gives you a wide range of different social and cultural context within that academic environment, where spoken English is used for academic purposes.

The talk that I’m going to give this morning is based on the Cambridge International Corpus, which is, I think, the largest in the world now. It’s got a billion words and it contains examples of both British English and American English. CANCODE, Cambridge and Nottingham Corpus of Discourse in English, which is a 5 million word corpus of spoken English, which I have worked on a lot in the last 10 years or so, various examples of business English and, again, a dedicated academic corpus.

There are dangers, and you can already begin to see them, I think, in this massive investment in describing English and the commercial exploitation of corpora,

but this is something I don’t want to necessarily deal with here, but we could take it up in discussion and questions later this afternoon.

What does a corpus enable you to do? Well, as I have already hinted, it gives you very precise information about the most frequent words in the language. Crucially for our purposes here this morning, it gives us information about differences between speech and writing, and because with a couple of clicks with the mouse, you can go from the wordlists to the precise sociolinguistic context in which the language is used, you can learn about different ages of speakers, different genders, different educational backgrounds, different regions, different language speakers and so on. The best corpora are sociolinguistically profiled in that kind of way.

Spoken corpora have developed fastest in the last 5 to 10 years. They are very difficult to put together. Each of you could put a written corpus together within a week, a million words, not difficult, just a case of scanning files of various kinds into Word, downloading some free software, and all of a sudden you can start searching for the particular words in your corpus. You can buy the whole of the *Times* of London for a year if you are so disposed and search within it for the most frequent words within that newspaper for a whole year and it’s something like 35 million words. It won’t surprise you to learn that “Blair” comes out in the top 10 most frequent words in the English language. So one of the dangers of corpora is that you have to make sure that your corpus is balanced, culturally, socially, in terms of different linguistic profiles and the like.

Here are – just as a matter of interest – the top 40 words from the CANCODE corpus. This is the sub-section, the 5 million word sub-section of the Cambridge International Corpus of English. If I had put up the written corpus, you wouldn’t have been that surprised. It’s mainly grammatical words, as you’d expect, the words, which provide the kind of glue that holds a language together. But, when you put up the top 40 spoken words, the language changes before your eyes kaleidoscopically. And it’s very interesting, I have been very surprised by some of the..., what the computer tells you, and it can print this list out in 30 seconds, computers can move that fast these days, and the software enables you to do these kinds of searches.

There is a mistake in number 26 – it doesn’t mean that we go round all day saying to each other “laughs, laughs, laughs”. It’s actually a stage direction. But it does illustrate how frequently we laugh when we are face to face with someone. It’s got something to do with the face-to-face nature of the encounter. There are various kinds of laugh, of course, but it is very significant in day-to-day communication. I have yet to see a course book or

textbook for teaching spoken English, where anyone ever laughs. It is deadly serious stuff. Deadly serious.

And here is the word ‘yeah’. I have yet to see any course book where people say “yeah”. Now, “yeah” is a top ten word in spoken English. Most course books have people say “yes”. We don’t actually say “yes” that often. In fact we say “no” more often, as you can see from item number 25. But “yeah” is very common. It’s got something to do with informality. It’s got something to do with the fact that “yeah” can also be a tag question – “we’re meeting at 6 o’clock tonight, yeah?” – that type of thing, rather than “aren’t we?”, which is what a lot of course books would have. I also think it’s significant that “yeah” is used by listeners rather than speakers. And we don’t teach listening much, well, we do, but we don’t always necessarily teach the vocabulary that goes with listening.

1	The	21	On
2	I	22	Oh
3	And	23	We
4	You	24	Have
5	It	25	No
6	To	26	Laughs
7	A	27	Well
8	Yeah	28	Like
9	That	29	What
10	Of	30	Do
11	In	31	Right
12	Was	32	Just
13	It’s	33	He
14	Know	34	For
15	Mm	35	Erm
16	Is	36	Be
17	Er	37	This
18	But	38	All
19	So	39	There
20	they	40	got

The top 40 spoken words from the CANCODE corpus

And one of the reasons why “yeah” is a top ten word here, is that it is used very commonly to back-channel to a speaker, that we are with them and we are following. Lots of the words on here are actually back-channel words - number 31 - is “right”. It’s got not much to do with right or left or with right or wrong, those are not the frequent meanings. It’s not in the top 40 because of those meanings. It’s in the top 40 because it’s a word to check understanding – “right?” – “got that?” and it’s a word to signal listenership, active, involved listenership, a very, very important part of two-way communication. And yet it doesn’t have propositional content, and that’s one of the problems with a lot of the words on this list, that makes them difficult to teach. “Right” here has pragmatic meaning rather than lexical meaning.

Number 15 – “uhm” – has pragmatic meaning rather than lexical meaning, as does number 17 – “er”. Are these words and am I right to put them in a top 40 list of the most frequent words in spoken English? The same would apply to other languages, of course. Are they words? Would you find them in a dictionary? If you wouldn’t find them in a dictionary, why not? Why would we not find the most frequent words in the language in a dictionary with a description of how they are used? What happens if you don’t use them as a learner of a language, any language? What happens if you don’t use them?

Some of them are particularly interesting – number 19 surprised me if you’d have asked me would that have been in the top 100 words, 500 words in spoken English, I would have said not. We know that “so” is used for purposes of intensification – “I was so cross” and “I was so upset”, you know, that kind of meaning. But the most frequent use of “so” is as a discourse marker. When I use “so” it’s often accompanied by a very irritating habit of clapping my hands. “So, we’ve covered ...” sort of thing. It is irritating, but what it’s doing is, it’s actually marking that you’re making a transition from one part of the talk to another. “So, we’ve done the 19th century, now we’re going to deal with the 20th century”, that type of structure. So it works as a discourse marker, it functions as a signal that you have changed topic. It functions as a kind of punctuation. Now, you can’t have punctuation in speech, many of these words are almost like punctuation marks, they are the equivalent of commas, semicolons, full stops, exclamation marks, questions, right? Spoken language can’t have punctuation, but it must have ways of signalling to the listener progression through your discourse. And they are crucial for all aspects of communication, which is why so many of them are here in the top 40.

Other words that surprised me include number 32, “just”, which I wouldn’t have expected to see there. And it has hardly anything to do with notions of justice. Think of the differences between “could you open the door?” and

“could you just open the door?”. What does the “just” do? Why do speakers of English, be they native speakers or successful users of English, why do they so frequently use “just” in everyday communications so that it is ranked so highly? What is it about “just”? what does it do?

It seems to lessen the imposition on someone, you are softening a request. If it goes with an imperative, you actually do the opposite: “just do it!”, “just shut up!” But used as part of questions, it lessens the imposition and is therefore a key component in politeness, at least as it is socially constructed and culturally constructed within this particular language. It would be interesting to see how far that also operates in other languages with similar words. Key word – very important – do we teach it? Do we teach it early in the curriculum? If we don’t, why don’t we? Are there reasons for it? What about many of these words, how many of them get into the early parts of a spoken language curriculum, in whatever language? But I’m talking here about English. What are the reasons for this?

I will have a little bet with you that number 28, “like”, will be in the top 10 if you were to invite me back in 10 years’ time. It’s one of those chart risers, you know, it’s going up at a hell of a rate, because, I think, or the corpus tells us, it’s because speakers under the age of 35, which discounts certainly me, but not many of you, of course, discounts people over the age of 35 from using it, it has the function of “said”. So, “my mum’s like, ‘make sure you’re back here at half past ten.’ And I’m like, ‘half past ten?’ And she’s like, ‘yeah, half past ten, you were late last night.’ And I’m like, ‘no, I wasn’t!’ And she’s like, ‘yes, you were!’”

That type of structure, which is very simple and very functional and very clear and very often the simplest words are the ones that get adopted. Look how many of these are simple in structure. Look how all of them are monosyllabic. “Like” has this function and therefore will very probably increase in that function and as old lags like me die, so the language will advance and “like” could well become Standard English. At the moment it’s not Standard English, because it is restricted to a particular age group. But processes of standardisation in the language, driven very quickly by these kinds of words, are something else I want to deal with as we go through things this morning.

OK, we can come back to these lists subsequently. Let’s just have a look and remind ourselves, with an example of the corpus, of what we do when we speak. And I would suggest this has the same functions in all

languages. Here’s a little bit of data from the corpus. It’s people talking in a vet’s about surgery:

<S1> *Let’s see ... we’ve already spent fifty for him and I want him to spend another hundred*
<S2> *Well*
<S1> *But that’s better than pins*
<S2> *Right*
<S1> *And surgery*
<S2> *Which would be another two hundred or*
<S1> *Yeah it’s more for a surgery*

Just have a look at the structure, because it is so very different from writing. There’s those words again, single words, often not containing any propositional content, so carrying pragmatic meaning rather than lexical meaning. Very, very, very difficult to teach therefore.

It’s easier to teach words that have lexical meaning, because you can point referentially to something. But words that have pragmatic meaning, you need almost the whole cultural context to explain it. And very often, intonation plays an important part. If you get the intonation wrong, you can very often get the pragmatic meaning wrong. You don’t have this kind of problem teaching written language. “Right” can be said in so many different ways, different intonation contours, different prosodies, carry different meaning, so it’s a very real challenge from the point of view of teaching.

Look at these so-called sentences. How far are they sentences? What happens when speaker one says: “and surgery”? So the sentences are built up, the constructions are built up across speaking turns, not as individual utterances. You hardly ever see this type of structure in course books for teaching conversation in any language that I’ve seen. Why? Well, I don’t think we had a lot of evidence for example of actually how spoken language works, so our mindset has been very much conditioned by our knowledge of the written language, which we are very good at and do know a lot about. But it dangerously preconditions how we describe and therefore maybe even how we teach spoken communication.

And there’s another one: speaker two, in their third utterance produces a relative clause. But it’s not relative to anything. A dependent clause that’s not dependent on anything. We produce this trick of “which” which is a good reason for doing it, very, very frequently in the course of a day, the corpus tells us.

What about “or”, what’s that doing? What is the speaker doing when he says “or” and then stops? Inviting the other speaker to come in? Signalling uncertainty? Opening up questions? Very difficult to decide. But if we are to teach – or at least examine the options of teaching –

spoken English more effectively, these are some of the descriptive questions we need to be alert to.

So in spoken language, it's probably fair to say that writers orient to norms and we know very well what a lot of those norms are. Most of the grammars and dictionaries of the English language presume written sentences. Most contemporary grammars of English illustrate the language by means of single sentences. So it biases our understandings towards written language, because speakers orient to each other, not to any set of norms.

Speech also takes place in real time, which is another reason why it is so difficult to do, even though we are very good at doing it. And it's the reason certainly why it is so difficult to teach. You can't easily go back and correct yourself, you can't delete and start again. It's in real time, not in the detached, decontextualised time that most writing involves. There aren't sentences in spoken language, well, there are in some versions of spoken language, for example giving lectures or reading the news, where things are written to be spoken, but very generally we don't speak in sentences or we sound very odd if we do. Quite a number of the utterances are incomplete – though is “incomplete” the right way of describing them? Because it sounds very negative whereas I think that a person's concern for the animal, expressed with “or” which is incomplete is actually very effective communication. We saw from the illustration in the veterinary surgery too, that language is jointly produced, it is co-constructed. That isn't the case in writing. And that the structures are very flexible. You can have dependent clauses and independent clauses operating separately from each other, which you can't have in writing. And small words are big words – just remember those 40 top words, and even the top 500 contain many, many small, single monosyllabic words, which are often the most difficult to get right in the classroom and which can also lead to difficulties in communication outside the classroom, because pragmatic meanings are misjudged or misinterpreted or misunderstood.

Look at the difficulties there are in teaching words like “well”, “right”, “just”, “at all”. As one of the checkout assistants in Sainsbury's said to me recently: “Have you got a reward card at all?” What is she saying there, why “at all”? “Have you got a reward card at all?” What does that mean? I still don't know. It's got nothing to do with “I'm not at all interested” – it's not that sort of thing, it's got something to do with politeness, maybe it's got something to do with routine, maybe it's got something to do with deference, I'm not sure. There's

a lot we don't understand in everyday communication, yet if she'd had missed it out, maybe it would have sounded less..., it would have sounded too direct, too assertive, I'm not sure.

“Sort of” is even more important and “I mean” is probably even more important, because it enables you to stop and start again. “I mean I should have said that”. So, complexities and difficulties, which are now made available to us, because corpora are showing and giving us examples through real everyday recordings in everyday circumstances. And this is not, you know, a criticism of previous descriptions of English, of course not. To get this type of data, only recent developments in technology have made it possible. To get this type of evidence of how language is used in everyday situations you need very, very good microphones, often microphones that can't be seen, because if people see the microphone, they start talking like they write. That's another interesting feature as well, the more self-conscious we become, the greater the formality level. So there's a lot of surreptitious, illegal, unethical recording that goes on.

And given that most of us spend 60% of our time talking about people who aren't present in a conversation, you can see that getting permission, ethical permission to use data is potentially problematic. You have to anonymise an awful lot of the data. I'm exaggerating a little bit, of course, there are lots of occasions where you genuinely, and must for good ethical reasons, tell people you're recording them in advance. But then we had to discount the first 10 minutes or 15 minutes of a conversation, because they are very conscious of the microphone. We've now got microphones the same size as my shirt button. So watch out – because with a direction finder I can hear what people are saying at the back of the room, developed by the CIA in the 1980s, now available to linguistic researchers in the 21st century!

The other thing that has surprised me in doing searches in spoken language corpora, is that of course words are significant and I've given you some of them, but so are chunks. In other words, of course, we do speak in words, but we also speak in sequences of words. And when you do frequency counts, you find they really shocked me as someone who's been in English language teaching for a long time, and foreign language teaching for a long time, how common the chunks are – in spoken English, not in written English, in spoken English. So, you know, words that we would teach like “fun” and “early” and “small” and the little sequences like “at the moment”, are almost equivalent in frequency to chunks like “and things like that”, “you know what I mean”, “a couple of”, often said very quickly as well. “Know what I mean?”, “know what I mean?”. It changes if you say “you **know** what I mean!”

So, can you see that the prosodic difficulties in teaching because... it's the same with "just". You can't say "Could you just open the window?", it changes the dynamic of the interaction completely. You've got to almost say it without anyone hearing it. "Can you just open the window?" Very hard to teach in a world where spoken English is all about enunciation and clarity and yet real language isn't altogether like that. And so there are very real challenges emerging in classroom terms and in curricula terms and in specifying as curriculum content, if we believe it is important as curriculum content – and we may not, we may decide that kids pick this up anyway. What we will need to focus on is writing development. But what we do need, I'll argue this, is at least some awareness of what the options are and that enables us, therefore, to operate comparatively in a more secure way.

But that interests me, that "that" and "things like that" is just as frequent as the word "small" or almost as frequent as the word "small" in that 5 million word spoken corpus. And not far behind it is "stuff like that". "Stuff" is the second most frequent noun in our corpus. Do you teach it?

Should you teach it? Is it necessary to teach it? What do you do if you use it? And people are very good at using "stuff". And if you say to someone, you know, "what have you got on this evening?", "Ah, I've got a bit of marking to do and stuff like that." That is probably as much as you'll want to hear. So "stuff" is deliberately vague. And the good communicators are those who are deliberately and purposefully vague. You don't keep many friends if you give them a list of all the things you are going to do in the evening. And they actually don't want to hear it. Unless you signal it as being particularly significant, you don't want a catalogue of all the things you're going to be doing that evening. "Stuff like that" is usually enough. And that's why we use it, that's why it's so frequent. Do you see it in textbooks? I don't see it in many textbooks, I don't see it in any textbooks. Because the textbooks are still assuming that the language is a written language, not a spoken language. That's the problem still.

Also very common, and perhaps less of a worry, but very common is ellipsis, which is by far and away the most frequent grammatical form in the corpus, and we don't need to dwell on this, it's just simply dropping noun subjects, often pronoun subjects and copulate verbs. But it does mean that the conversation goes along at the right level of formality if you use ellipsis. So in this example

<S1> *You're going to Leeds this weekend?*

<S2> *Yes I must*

S2's response is fine, but a lot of course books to teach you spoken English have people reply and converse in complete sentences. Very odd. "Are you going to Leeds this weekend?", "Yes, I'm going to Leeds this weekend." "Are you staying in a hotel?", "Yes, I'm staying in a hotel." "Is the hotel in the centre of the city?", "Yes, the hotel is in the centre of the city." It sounds very menacing, you wouldn't want to meet the person.

I am being a little bit facetious, of course. In the early stages of language learning, you need those sentences, don't you, because it's a way of practising structures. But there comes a point – and this is what's difficult to judge in pedagogic terms – there comes a point where the tension between full and less full, between complete and less complete is a difficult pedagogic decision to make.

Here are some other features, which do not appear in conventional grammars, because we don't write them.

"She's a very good swimmer, Jenny is."

"It's difficult to eat, isn't it, spaghetti."

"We're gonna have steak and fries, we are."

"It can leave you feeling very weak, it can though, apparently, shingles, can't it?"

Would you write this? No.. Is it correct? Yes. Is it Standard English? Yes. But no grammars of Standard English give evidence of this as being standard. Therefore the assumption is, if you're not careful, that it's non-standard. Because standardisation, in a lot of the examples of English at least and, I would argue, other languages, is based on written norms, not spoken norms. And you need both.

If you were writing, you would have to write "Spaghetti is difficult to eat." In speaking, I think you are showing a great deal of communicative sensitivity to your listeners, because you're repeating, or reintroducing, the noun that links with the pronoun. Just to check that they've understood it. So it's a very subtle strategy and essentially interpersonal in orientation. So much of language that we teach, I think, is, quite properly, transactional and transmissive because it's information-based. But we also have to learn to balance that with more interpersonal, interactive communicative features, and this is a sort of example of interpersonal grammar. Grammar used for rapport between speakers. And all languages will have structures like this because they're necessary. You can't just simply have conversations with people that involve information transfer. Language is not about information transfer. It's also about building interpersonal, relational discourse.

The other thing it shows you is that the language is changing and the corpus shows you that a lot of speakers of English have started, in spoken discourse, to remove distinctions between singular and plural. So “there’s lots of cars in the car park”, I would argue, is now standard in speech. If you are writing, “there are lots of cars in the car park”. It’s the same with deictic uses of “there”. You find this in public discourse as well. There are 15 notices of public figures speaking that we recorded in the corpus,

“What happens is that there are 15 members of the Security Council”, correct, “There’s 5 permanent members”, incorrect, and “5 permanent members who got the veto”.

Not incorrect, it depends on..., it shows..., I suppose, it shows you language in transition.

These guys are responsible for an awful lot of mistakes. The one who said it is on the left. The one on the right can’t speak English at all.



The other set of issues for spoken language communication is that the place of English in the world is changing very, very fast indeed and we do know that with the investment by the Chinese in primary English we will see an astonishing shift towards English used by speakers as an additional, second or foreign language by 2020. These figures are from a fascinating book, which I’d recommend to anyone, available on a free download from the British Council, called “English Next” by David Graddol. It’s in the references that I’ve given to you. Do read it, it’s a brilliant prospective survey of where the English language is going in the world and its conclusions are that..., you know, the figures just present very real problems because it will see a massive increase in the use of the English as a lingua franca, that is a language for international communication between speakers for whom English is not a first language. And David Graddol predicts that by this time 80% of communication through English will be between speakers for whom English is not a first language. I don’t like native/non-native

distinction, it’s a very awkward, false one, but he says between non-native speaker and non-native speaker.

This has, of course, massive implications, I think, for the EAL community and teaching EAL. It raises problems for the teaching of English. If we continue to assume that the only speakers we produce are literate speakers, i.e. speakers who speak like they write. There are problems too with a corpus that is only a verbal corpus – that’s one of the big weaknesses, isn’t it, at the moment of all corpora, they’re not visual. And as we know so much of communication, face-to-face communication, involves different and culturally differentiated forms of proxemics, kinesics, eye contact, gaze, body language and so on, all of which have significant cultural implications, all of which are necessarily embedded in the choice of particular linguistic forms.

You can see why people concentrate on writing. First it has prestige, literacy is associated with speaking and writing and all curricula push, I think mistakenly, all they push is reading and writing. And I’m not, of course, denying, that isn’t highly significant economically and socially and in terms of getting jobs and so on, but if it’s at the expense of understanding better and helping people become better at spoken communication, then there are also prices to pay.

It raises, too, the problems of how we define fluency and there are in a lot of curriculum documents, I think, rather dangerous and misleading statements about fluency. There is the assumption that fluency, certainly in a lot of the National Curriculum documents, that fluency involves using Standard English. Well, since we are problematising what Standard English is, I’d like anyone to tell me what they think Standard English is, or, at least, I’d ask whether some of the features and forms that we’ve looked at count as Standard English or not, because according to the curriculum statements they don’t. But for me, and I hope also for you, they are singularly important as components in successful communication and are therefore examples of standard spoken English.

“Smooth, continuous talk” is used a lot, “flow” is used a lot and “grammatically accurate” is used a lot. They really do want people to sound like newsreaders when they speak or like members of a debating society. The consequence, therefore, is that dysfluency gets constructed as hesitant, sloppy, people who can’t remember words, people who repeat themselves and code-switch between languages.

Is this, where there is a lot of repetition, fluent or not?

“Jim, did you...did you...did you see David at the meeting last night? No, the night before, wasn’t it?”

Is that fluent or dysfluent? According to a lot of National Curriculum documents it would be dysfluent. I would argue that that shows – and if you had the visual accompaniment to it you'd see – that the speaker paused because the person they were talking to was doing something else, stopped, paused, rephrased, restructured, repeated, forgot what he was going to say, recovered his ground, corrected himself and got the question right at the end. In the second example, which has an example of listenership built in, therefore an example of confluence rather than fluency. Because I think fluency is not singular, fluency is plural, things need to be co-constructed, co-developed, shared, managed jointly.

<S1> *So what did marketing do?*

<S2> *They did it that way and they introduced, yeah, yeah, and last year they introduced 8 new products in just six months, 8 is huge, it is, isn't it, know what I mean?*

Sloppy? Careless? Hesitant? Full of grammatical errors? Or successful communication? Where repetition can be used to plan strategically, to be sensitive to the turns that we all take in conversation, can have appropriate levels of vagueness in them, to invite the listener into the discourse, “know what I mean?”, can clarify and confirm, can summarise, can hold onto the floor where necessary and can use repetition and many of those other forms for emphasis. So the language is not just organisational and transactional, which is as a lot of curriculum statements tend to assume it to be, but also interpersonal and relational.

Big, big, big problem: how do you test this? And it's one of the reasons, I think, why we have all struggled to get spoken language more fully into the curriculum, is that the best testing mechanisms are those for testing single literate speakers speaking grammatical, accurate English. A lot of work needs to be done, maybe with video support to develop understandings that this type of discourse is not only normal and natural and what many people do in all languages during the course of a day many, many times, but also successful communication. Repeating yourself is a sign of success in a lot of contexts. Pausing is a sign of success. Hesitation, because it's collaborative, is also an index of success as well. So there are really very, very big and serious problems in assessing and developing mechanisms for testing this type of competence.

Just a few minutes left to move onto one final statement, because I have, I think, rightly or wrongly, suggested a division between spoken English and written English, and it's not quite as simple as that, as

you all know. With the advent of texting and email and so on, a lot of electronic communication has us write as if we are speaking. Largely because we are doing it quickly, we are doing it in real time, largely because we are all getting better and quicker at keyboard skills, largely because some online communication is in real time or more or less in real time. And therefore there is this now overlap between spoken and written English. And also in our corpus we've got some very rich samples of not first language speakers of English communicating in a mix of English and other languages. This is one from our corpus of Cantonese students at the university writing to each other. We've got a million words of this. Very, very interesting development: they could download into their email systems Chinese characters. They chose not to, they use English, they want to use English to communicate with each other, but English is insufficient, so they mix not only text language but also transliterated Cantonese into their text messages. Forgive me for poor pronunciation here, but it reads something like this:

<S1> *It's snowing quite strongly outside, be careful.*

<S2> *I will, thanks.*

<S1> *We will mhm. Dim, er.*

<S2> *OK, la, juz got back from Amsterdam, lo, how are you.*

<S1> *OK, la, I have nine tomorrow.”*

[That's a lecture at nine.]

<S2> *Ha ha ha, I have 2 to 4. Soooo happy. Che, anyway, have your rash gone?*

<S1> *Yeah, but I have scar, oh ho, ugly, ah.*

<S2> *I see, I sigh, I see.*

<S1> *Nga, er, la.*

<S2> *Still a pretty girl, ha ha.*

And there were translations from Cantonese there – now might not this be one version of a future for inter-cultural, inter-lingual communications through English with layers of identity as Cantonese speakers preserved within the discourse with all these layerings of speech and writing interfaced with one another? We're still studying this, it's a very interesting phenomenon and common across many languages where English is the medium of emailing or texting or whatever.

I don't need to read the first one to you, it's an example of an email where it contains many of the features we've been discussing as common to spoken standard English, but it's written. Ellipsis is a good example. And the last one I will read to you because it shows that the *Daily Telegraph* can't use Standard English either. It's from a journalist, a very good journalist, writing for the *Daily Telegraph* who is interviewing Mick Jagger and writing up the report, using very interactive discourse markers to involve and engage the audience. It's a very creative piece of writing, but an example with all these discourse markers and many of the words you saw on the list at the

beginning of the talk recycled there, along with ellipses, to produce a kind of hybrid between speech and writing.

“Right, so there I was, sitting in Mick Jagger’s kitchen while he went about making us both afternoon tea. Well – you can imagine how long it took to get him to talk about the band’s latest album. Exactly, you’ve got it...over 2 minutes.”

So, written English but containing many of the features of spoken English and formal for a *Daily Telegraph* magazine, but also surprisingly informal and interactive and interpersonal.

OK, what might we summarise from this? Well, modern corpora of English tend to show us things we didn’t know before, tell us things about ellipsis tails, flexible clause structures, vagueness in language, that do not occur regularly in written English but which are a common, and I would argue, standard feature of spoken English. These forms were often unnoticed previously because we haven’t had recording technologies supply us with the information. You need a new meta-language to describe them – “tails”, for example, is not a standard grammatical term at all. “Vagueness” does not appear in a lot of standard grammars. Pity, because, I think, vagueness is really very important for communication. Some of the structures overlap, speech and writing, and we shouldn’t ignore the overlaps and continua between speech and writing. Fluency itself has been under-theorised. The teaching of English still works, or the teaching of spoken English still works from assumptions based on written language norms. And we may think this is OK, I’m quite happy to hear arguments that in spite of all the evidence I’ve shown you this morning, we still do, as a community, believe that focussing on writing is the single most important thing. All I would argue is that we do need to know and think about and take on board these questions to do with spoken communication and how we describe it, how we theorise it, what impact it has on standardisation and our understanding of standard forms. If we are then to reject it, we reject it in a very principled way.

Spoken language focuses on speakers and listeners, listeners are particularly important. Listeners are not passive, they are active and they co-create with speakers. Fluency, therefore, is confluence, not simply fluency.

I would argue finally that there are dangers, aren’t there, in these investments in large, million word corpora that you create a hegemony, if you are not careful, of corpora that only look at a single language.

And I think that is especially dangerous in an EAL world, but it’s a starting point and it raises questions which I do believe are pertinent for all languages and I look forward to you discussing and hearing your questions about this later on in the afternoon.

Thanks very much.

References:

- Biber, D. et al. (1999) *The Longman Grammar of Spoken and Written English* (Longman, Harlow)
- Carter, R. and McCarthy, M. (2006) *Cambridge English Grammar: A Comprehensive Guide to Spoken and Written Grammar and Usage* (CUP: Cambridge)
- Cornbleet, S. and Carter, R. (2001) *The Language of Speech and Writing* (Routledge, London)
- Halliday, M.A.K. (1989) *Spoken and Written Language* (OUP Oxford).
- Pridham, F. (2001) *The Language of Conversation* (Routledge, London)