

THE LONGMAN SPOKEN AMERICAN CORPUS: providing an in-depth analysis of everyday English

Karen Stern

Providing good coverage of American English in Longman's dictionaries from this side of the Atlantic might seem a difficult task. How does a Yankee over here keep from falling victim to the insidious transatlantic drift? The Longman Spoken American Corpus, owned by Pearson Education and gathered by Professor Jack Du Bois and his team at the University of California at Santa Barbara (UCSB), is one of the most important tools our American team of lexicographers has for presenting students with American English as it is used every day. The spoken corpus, used in concert with two corpora of written American English, gives us full coverage of the language in all its forms.

The Spoken American Corpus is a five million word database gathered from 12 regions across the continental US. Equal numbers of participants were chosen from each region, and a balance was struck between the numbers of participants from rural and city areas within those regions. UCSB sent project workers out to each of the regions to deliver portable tape recorders to the participants who then recorded four hour chunks of their normal daily conversations over periods of at least four days. Records were kept of the situations being recorded, and of the demographic details of everyone involved in the conversations.

The conversations were recorded as unobtrusively as possible, with the tape recorders simply being allowed to run for four hours at a time wherever the conversations were taking place. The tapes were edited to weed out silences and long stretches of garbled material and, finally, transcribed for Pearson Education team of American keyboarders. None of the last names, addresses, or telephone numbers that were talked about on the tapes were transcribed. This combined with the fact that participants were guaranteed anonymity resulted in the natural conversations we required.

The participants were chosen to be representative for gender, age, ethnicity, and educational experience. The demographic breakdown follows the latest US census statistics:

GENDER:	Male 50%	Female 50%			
AGE:	*18 24 20%	25 34 20%	35 44 20%	45 60 20%	60+ 20%
ETHNICITY:	White 75%	Black 13%	Hispanic 8%	Asian 4%	
EDUCATION:	Degree/Higher Degree: 33%		College 33%	High School 33%	



www.longman.com/dictionaries

LONGMAN DICTIONARIES
Express yourself
with confidence!

(*children are also represented in a large number of conversations throughout the corpus)

The Spoken American Corpus is used, as all our corpora are, to provide the student with grammatical, lexical and frequency information as well as with natural examples of typical usage. As several people in these pages have attested, spoken language yields a number of linguistic features that have only recently begun to be covered in learners' dictionaries. Although one or two of the items of lexis, grammar, and discourse that follow have begun to be used in Britain, they are all uniquely American in origin.

SLANG

For the student who likes to be up on the latest slang, the Spoken American Corpus presents us with, among other things, **buff**, **fine**, and **way to go** all terms of praise and general approval used by the younger generations. A typical corpus line for **buff**, used to describe someone with a particularly well formed, well toned physique follows (taken from a conversation in which a young woman describes a former college classmate):

Example: ...you know he was like a total skater like cute little buff guy freshman year, but he was cute at Davis still too

Fine is also a term of admiration for an attractive person, although it does not specifically describe the body, and might better be defined as "very good looking". A case in point is the simple and deeply felt utterance of one man on the corpus: *A bunch of fine ladies out there.* The same women who knew the "little buff guy" mentioned above, also use **fine** in the following lines to describe another one of their classmates.

Example: I remember all the guys saw this picture and they totally wanted to meet her they were like God that Ursula girl is fine, then they finally realized that she wasn't all that.

Evidence from the Corpus also suggests that **fine** can be used to describe things or places as well as people, as in a conversation about a vacation in Tucson, Arizona that describes various activities in that area as "so fine".

Way to go is probably the oldest of the three, having started its life over 20 years ago as a cheer used in sports events by cheerleaders to urge the team on. Its status has been upgraded since then to that of a general statement of approval for something that someone has done well. It is surprising to find that none of the major American dictionaries' latest editions cover this phrase, especially in light of the fact that it is now frequently used sarcastically a sure sign that a phrase has found its place in the language. Here, two women are talking as one is cleaning house:

Example:	
Speaker 1:	Crash.
Speaker 2:	What did you break?
Speaker 1:	I broke my <unclear> holder.
Speaker 2:	That teal one?
Speaker 1:	Yep
Speaker 2:	Way to go.

NEW USES FOR COMMON WORDS

As well as pointing out new words that are only used in spoken English, our aim is to give students useful information about the way words are used. The Spoken American Corpus has given us some interesting insights into new senses of very basic words that students may well assume they have already mastered. So, for example, students hearing *Boy*, that is one ugly car might well wonder where the other one is. To clear up any confusion, the forthcoming Longman Dictionary of American English (LDAE) gives them the following functional definition and example at **one**:

spoken said in order to emphasize your description of someone or something: That is one cute kid!

To a student who has been taught that we use *easy* to refer to things that we learn or work at, the phrase **I'm easy**, meaning "it doesn't matter to me what we do", might cause consternation. Many of the learners dictionaries cover the phrase, which has become popular on both sides of the Atlantic, but, interestingly enough, the 10 million word spoken element of the British National Corpus (BNC)' does not contain any examples of the phrase. The Longman Spoken American Corpus, however, has several lines illustrating this use. For example:

Example:	
Speaker 1:	Mhm, So I wonder what I should do, shall I do that, leave my stuff here and then come back here? But you want to go into the mall, right?
Speaker 2:	I, I mean, I'm easy. I can do whatever you need to do. I can just take you out there and bring you back or..

GRAMMAR

The corpus is providing us with the long awaited evidence we need to further the cause of spoken English in the classroom. McCarthy and Carter in their article *Spoken Grammar: What is it and how can we teach it? [2]*: point out that the spoken language has its own grammar and that "learners" need to be given choices between written and spoken grammars". This fact has only just begun to change in Britain in the last few years with the advent of spoken corpora. In America, the spoken revolution has yet to take root.

Take for example the phrase **you guys** which has been in use in American English for at least 40 years. While most American dictionaries give a basic definition of this sense of **guys**, very few of them explain how it is actually used. The word indicates not just the two or more men to whom you are speaking, but any two people male or female. It has come to be used (along with "y'all" in the southern states) as a substitute for the plural form of "you".

Over half the corpus for guys is taken up with the phrase you **guys**, as in the sentence *Hey, are you guys ready to eat?* However, another eighth of the corpus material shows that these guys and those guys are also used commonly when the speaker is talking to or about two or more people without reference to gender. A few cases in point:

Example 1:

... It seems like those guys change their divorce agreement almost [weekly]

Example 2:

But, uh, those guys would kill me if I didn't show up to play at this party.

Example 3:

...we were gonna play Trivial Pursuit and couldn't decide how to divide up the team, so I said alright you three guys against me

While some might argue that this is not standard English, this use of **guys** occurs 600 times per million words in the corpus, making it far too frequent to ignore. Certainly, students wishing to be able to speak real American English will benefit from seeing how it is used. The new LDAE gives the following definition:

you guys/those guys *spoken* said when talking to or about two or more people:

*We'll see you guys Sunday, okay?
Are those guys coming with us or not?*

DISCOURSE

Attempts to mimic American English often involve beginning sentences with **I guess** and inserting **like** before every third word, and although using these terms does not guarantee a successful imitation, the corpus confirms that they are indeed amongst the most frequent utterances in spoken American English (**I guess** occurs 756 times per million and **like** occurs a staggering 9,010 per million). More interesting, however, are some of their more subtle uses in discourse.

We already know that **I guess** is typically used to show uncertainty and is often used in statements in which a conclusion is drawn, as in the sentence *He said he wasn't hungry, so I guess he ate at Ron's*. What the corpus shows is a use of the phrase that is neither of these. Looking at the following lines we see it being used to report events that the speaker knows about secondhand.

Example 1:	
Speaker 1:	Who yelled at you?
Speaker 2:	Well, Greg didn't yell, but I guess Mr. [name] really had a fit about it.
Example 2:	
Speaker 1:	She was a nanny.
Speaker 2:	A nanny. Yeah. But then I guess she came back home...

The information reported after **I guess** is not known firsthand by the speaker in either of these cases. However, it is not uncertain, and unlike the statement *...so I guess he ate at Ron's*, none of these lines show that the speaker is trying to draw a conclusion from pieces of information he or she has been given. These are statements that are known to be true, but that the speaker introduces with **I guess** to show the listener that the knowledge is secondhand. This use is more frequent than one might expect, occurring approximately 70 times per million words.

Like is not always the meaningless filler of space that it might seem to be at first glance (although sometimes it is very difficult to ascribe a function to it). In the phrases below it is used with the auxiliary verb "be" (e.g. **I'm like**, **he's like**, etc.) to introduce reported speech.

Example 1:

He says, do you run? I said, yeah. He's like, how fast? So I'm like. oh.

Example 2:

...another song comes and she's like, you wanna go dance? and I'm just like...alright

In the lines above **like** is used as a type of substitute for the **verb** say, and is also very close in meaning to **go** which is used in both American and British English to report speech (as in: *Mum goes, don't you swear at her*).

Like has many other functions in spoken American English. In the following lines it shows shyness or a hesitancy in broaching delicate subjects:

Example 1:

Like could you afford to just give her like twenty five dollars a week or something instead of making her ask for it?

Example 2:

Like did you feel kind of emotional when he starts talking about that?

It is used to mean "for example" when someone begins to explain something or tell a story:

Example 1:

... like there's a technique you can use like if I were using erm, er the I, the first person singular...

Example 2:

...the only thing with the sushi in this town is so overpriced you know? Like I went to this stupid little clone or <unclear> it's like twenty bucks

The list of functions goes on, and given that, according to the corpus, Americans say like once in every 110 words, it is a word that warrants much more attention.

The Spoken American Corpus is the latest addition to the growing body of corpora, and it is the first database that documents American spoken English on a substantial scale. Its first applications can already be seen in the third edition of the Longman Dictionary of Contemporary English. It has also been used extensively in compiling the forthcoming LDAE. Given that there is always more slang to discover and that the possibilities in the areas of spoken grammar and discourse are seemingly endless, we are very glad at Pearson Education to have this tool to help us investigate the language further.

NOTES

- 1 The British National Corpus is a collaborative corpus collected by Longman, Oxford University Press, and Chambers, with the universities of Oxford and Lancaster. The BNC was partly funded by the British Government and is easily available for research within the European Union. Longman's particular contribution was the creation of the Spoken Corpus of the BNC.
- 2 McCarthy M, Carter R 1994 Spoken Grammar: what is it and how can we teach it? *ELT Journal*. July 1995: 207