

What research questions does the article address?

Erman's stated overall aim in this study was "to review the functions of pragmatic markers in spoken discourse" by drawing on examples from the Bergen Corpus of London Teenager Language (COLT) corpus and the London-Lund Corpus (LLC). She focuses especially on the use of "you know" as a marker in these domains, and contrasts adult and adolescent uses of this marker across these domains. The domains and functions are:

- 1) The textual domain, in which speakers structure and organize text, edit text, and 'repair' what they're saying as they go along.
- 2) The social domain, which is focused on the listener and is used as a "comprehension-securing function," or as a "turn-taking" device. In this domain, the speaker wants to make sure the listener "has correctly understood specific references made in the text, usually to people but also to objects and other phenomena," writes Erman.
- 3) The metalinguistic domain, which Erman proposes to accommodate discourse markers that focus "not on the text or the participants, but on the message as a whole." Erman tells us the function of the metalinguistic is modal and is "directed towards the speaker's subjective appreciation of the illocutionary force of the utterance as a whole." (I wonder if Erman is referring to epistemic modality?) According to Erman, discourse markers in the metalinguistic domain can enhance the effect of the message (illocutionary force) and emphasize the speaker's authority, and act as face-saving devices (hedges, approximating functions).

Why do you find the research question interesting?

I found this topic promising because I'm interested in how and why various expressions make their way into the common vocabulary. It's also interesting how innovative young speakers are with their language. For example, about 15 years ago (more or less?), the "I'm like" and "he's like" construction (rather than saying "he said" or "I thought or said...") started making its way into young people's talk. Today, that form is pervasive.

Also, some of the concepts Erman touched on (briefly) in the paper, especially the "chunking" concept and the possible connection to construction grammar, sounded interesting. I'm exploring constructions that can give rise to humorous wordplay (eg, the "waiter, what's that fly doing in my soup?" / "I believe it's the backstroke" example from Fillmore), so this section of the paper seemed interesting. However, these ideas were presented only briefly and not pursued (although Erman does suggest in her conclusion that "you know" may be on its way to being further pragmaticalized, and that its use as a chunk in prefabs provides some evidence of this.)

What conclusion about the question does the article draw?

Erman makes a few general conclusions about the differences between adolescent and adult speech with respect to discourse markers with focus on 'you know,' including:

- 1) Adults use 'you know' mostly in the textual domain, to organize the content (theme). Adolescents use this domain far less frequently, and when they do, it's not thematically.

2) Adolescents use ‘you know’ mostly in the social domain, especially as a “comprehension-securing” mechanism, but also in the metalinguistic domain. Erman concludes that adolescents use the “you know” as a discourse marker in “a modal, metalinguistic function” that mostly emphasizes “the illocutionary force of the utterance as a whole.” On the other hand, adult use in the social domain was considerably less, and use in the metalinguistic domain was virtually nil.

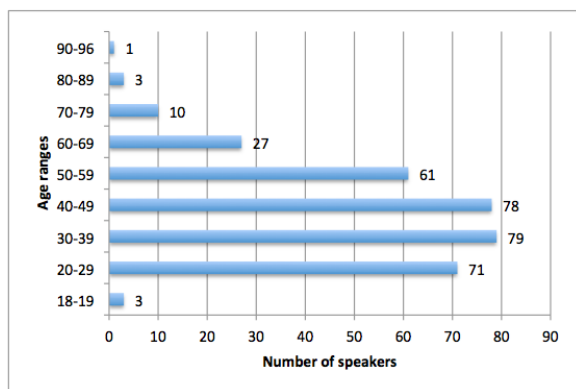
Erman also suggests in her conclusion that the functions in the metalinguistic domain and the use of “you know” as “chunks” in pre-fabricated speech may point to ‘you know’ “moving in the direction of being further pragmaticalized,” but that it’s premature to draw this conclusion.

Do you find the authors’ approach satisfactory? If not, how else would you do it?

The more time I spent looking at additional material (including the web links in your slides) to try to figure out the point of this study, the more disappointed I became with this paper. For one thing, the data Erman compares is drawn from what appear to be two completely different approaches to corpus creation. For example, the LLC contains both public discussions (“dialogue heard by an audience that does not participate”) and conversation. I was surprised to learn that “conversation” included “private conversation” both “face-to-face” and “telephone conversations,” some portion of which had been “surreptitiously recorded” (“ye olden days,” :-)) before wire-tapping laws?). I pulled some of the LLC speaker details into an Excel spreadsheet¹:

Age	Number of speakers	Age	Number of speakers	Age	Number of speakers	Age	Number of speakers	Age	Number of speakers	Age	Number of speakers	Age	Number of speakers	Age	Number of speakers	Age	Number of speakers	Age	Number of speakers
18	2	20	12	30	36	40	29	50	27	60	10	70	5	80	1	90	0		
19	1	21	3	31	1	41	3	51	1	61	2	71	0	81	0	91	0		
		22	3	32	3	42	4	52	7	62	2	72	1	82	0	92	0		
		23	3	33	3	43	6	53	6	63	2	73	0	83	0	93	0		
		24	3	34	2	44	8	54	5	64	3	74	0	84	0	94	0		
		25	22	35	12	45	13	55	8	65	4	75	1	85	0	95	0		
		26	4	36	9	46	2	56	3	66	2	76	1	86	0	96	1		
		27	2	37	2	47	1	57	1	67	1	77	1	87	2				
		28	7	38	8	48	6	58	3	68	0	78	1	88	0				
		29	12	39	3	49	6	59	0	69	1	79	0	89	0				

Here’s another view of this data, summarized by age group:



So here we have 330 speakers with an age span that runs from only a year or two older than the COLT speakers (18, 19) to many decades (and generations) older. (Although Erman does point out

¹ Not included is another 100 or so speakers (rough guess) whose ages were listed in the LLC data as non-specific ranges, e.g., “20-30,” or “20s,” so I think that the LLC represents close to if not more than 450-500 different speakers.

these differences, it still makes me wonder why she chose to use these two corpora in this way, other than perhaps because she had access to them?)

On the other hand, the COLT participants were a mere 31 boys and girls between the ages of 13-17, who evidently (per the COLT webpage) were “recruits”—so we can presume that these adolescents knew they were being recorded (which can influence their speech (discourse) to some extent, I’m guessing?), while we can also assume that some portion of the adults did not know they were being recorded (the “surreptitious” factor). Given these differences, plus the different timeframes (1960s, 1970s, etc) I really don’t believe that any conclusions can be drawn or theories tested using data from these two corpora.

In pursuing a study such as this, I would try to start such a study with common timeframe; or, if not common timeframe, at least draw data such that sample populations being compared were defined using the same criteria. So use only the data from LLC that has the same characteristics (surreptitious recording or not, for example, face-to-face or phone? and so on).

Furthermore, as you mentioned in class, Erman did the identification/labeling of the various instances of “you know” as to discourse domain and function (text structuring, text editing and correction in the textual domain, for example, or turn-taking in the social domain) by herself, so she was working in a vacuum. As you suggested, it would have been better if she had other researchers evaluating the texts, and at the very least, confirming the categories she assigned. She could have then factored in the number of researchers who came to the same conclusions as she did to weight or score the results before drawing conclusions. The numbers that she does provide might then be more meaningful.

* * *

Out of curiosity, I went to Mark Davies’ relatively new (Aug 2012) “Corpus of American Soap Operas,” a 100-million word corpus, with transcripts going back to 2001. A quick search through this corpus reveals that the pattern “, **you know**,”² occurs 64,083 times. Here are just a few samples from that corpus:

Chris: You know, um, when your mama died, more than once, **you know**, just walking down the street, seeing someone who reminded me of her...

Brooke: First of all, I want to thank you, **you know**, for all the help and support that you’ve given Laura and me these...

Liza: No. No, we don’t. But, **you know**, we just take it -- **you know**, you do what feels good and

Nurse: Oh, that’s just, **you know**, second nature. You gather them, bag them, give them to a

Ryan: I’m just saying it’s, **you know**, a little risky. Hey, but that’s your style. You’re

Rosa: I wanted to, **you know**, at first. But then he changed. He turned completely different and --

Dr. Clader Well, **you know**, the baby gets examined right after birth -- remember? -- And a few

Of course, in the realm of pragmatics, these should be looked at in the context of a conversation, to try to determine if they are indeed being used as discourse markers. But since this is scripted daytime drama, I’m not sure how that would be approached.

² Addition pattern results: “...well, you know...” (8,600); “...well you know...”(31); “...but, you know...”(5,030); “...but you know...”(1,632); “...just, you know...”(2,057); “... so, you know...”(844).

After looking at the Soap Opera corpus, I learned that the Corpus of Contemporary American English (COCA) also contains spoken text, so I took a cursory look at that corpus as well and found thousands of examples of the same pattern.

What's interesting (from an admittedly unscientific perspective) is that the COCA spoken sample corpus is only about 5-6% smaller than the Soap Opera corpus (~95 million to 100 million), yet the number of instances of the same pattern "**, you know, "** in the spoken section of the COCA corpus is almost 50% larger than that of the Soap Opera corpus. This was interesting to me mostly because I had thought the use of "you know" in the Soap Opera corpus might have been over-represented, since the dialogs are contrived, and yet, in actual speech, "you know" is really quite common:

	Corpus of American Soap Operas	Corpus of Contemporary American English (spoken)
Words in corpus	100,000,000	95,385,672
Total pattern " , you know, "	64,083	110,444

Here are just a few examples pulled from the speech-portion of the COCA corpus. The sources for the transcripts include broadcast news (ABC News, Nightline, CBS, Sixty Minutes, for example), and talk shows such as CNN Larry King, Fox Glenn Beck, Oprah...), as well as radio (NPR All Things Considered, Fresh Air, and so on).

...out about Judge Sotomayor did, **, you know,** reflect these rumors that **, you know,** she can be a bully on the bench and she can be tough on...

We were living together as **, you know,** a kind of, I guess, unmarried couple or whatever you want to...

hills out there, so the wind swirls around the hills, and **, you know,** we're coming around following the river and looking for things and then we

prefer that figure skating remain this wonderful thing that once every four years **, you know,** rises to the heights and it's something glorious to see.

our brains evolved in an era when there wasn't much need for **, you know,** mathematical computation, for a sharp sense of quantitative thinking. Why?

...have, like, a smile on your face and your shoulders back **, you know,** and your head up, you know, that kind of exudes confidence in

, when they put me on the stretcher. I tried my hardest **, you know,** put all my heart into it just to let them know I was all

Finally, here's an excerpt chosen from COCA (spoken) at random, a more complete exchange from a 2011 broadcast of from NPR's "Talk of the Nation," hosted by Neal Conan:

...not two had passed when the hurricane hit. So all of the sudden, I am, **, you know,** in a place - devastation like I've never seen. **You know,** there's mud 10 to 15 feet high, **, you know,** between the roads. There's rocks the sizes of cars that we're taking out of people's homes. And I have just never seen such devastation.

CONAN: And...

MICHAEL: And yet the people were so resilient and they were joking around, I mean, people who had lost absolutely everything **, you know,** people who didn't trust banks and had all their money in their homes.

CONAN: And...

MICHAEL: And they somehow had a good, **you know**, they somehow managed to have a really positive attitude. And at that time, I was really, sort of, feeling sorry for myself and it really, really put things in perspective.

CONAN: I was going to ask if that totally changed things from being one kind of an experience to completely another.

MICHAEL: Yeah. It's interesting.

What does all this prove? I'm not sure, but it seems obvious that "you know" is used quite often in adult American speech. It's quite possible that "you know" is really a "chunk" or possibly a construction, in and of itself, but to figure that out would require designing a better study using appropriate statistical techniques and sample design. Personally, I do think "you know" is often simply used as "filler," when people's brains get out of synch with their mouths in one way or another, and they are simply trying to piece their thoughts together into a meaningful utterance.