



1-1-2001

# Why You Can't Do a VARBRUL Study of Quotatives And What Such a Study Can Show Us

John Victor Singler  
john.singler@nyu.edu

---

# Why You Can't Do a VARBRUL Study of Quotatives And What Such a Study Can Show Us

# Why You Can't Do a VARBRUL Study of Quotatives And What Such a Study Can Show Us

John Victor Singler

## 1 Introduction

The twentieth century saw the introduction of three quotatives into American English.<sup>1</sup> First there was *go*, whose appearance appears to date at least as far back as the 1940's and 1950's, according to the recollection of those who were teenagers then. After *go* came *(be) like*, first noted in Butters (1982). In some parts of the United States, notably California, *(be) all* has now followed—and to some degree supplanted—*(be) like*. Three consecutive quotatives in the course of a narrative by a New York City female college student

---

<sup>1</sup>The data for this study were provided by Samantha Abrahamsohn, Nicole Abrams, John Allison, Bryan Anderson, Chris Anderson, Yuliya Angelovskaya, Heather Barlin, Marisa Bassi, Sharmila Basu, Frada Berenshteyn, Amanda Bernstein, Dana Bivacca, Ashley Blackwood, Melissa Boltax, Jill Boncek, Abigail Braithwaite, Cristina Buccola, Steve Carver, Natalie Charles, Pete Chatmon, Kimberly Cook, Laurie Crist, Karen Cummo, Kelly Cunningham, Niels Dachler, Lila Damavandi, Roslyn Dames, Larissa Davidov, Samantha DeCerce, Jay Diaz, Danielle DiTalia, Florence Emerole, Rachael Evans, Sawyer Fischer, Andrew Fleming, Michael Foster, Jessica Geminder, Greg Goldberg, Suzann Goldberg, Lauren Goodman, Amanda Hamann, Shari Headley, Dana Horoszewski, Francis Hult, Noriko Ishibashi, Aiko Ishikawa, Maria Isidorou, Orly Jalowski, Clement Joseph, Ericka Joseph, Jennifer Kaplan, Katherine Kearney, Allegra Ketchum, Christine Kim, Heedal Kim, Sun Kim, Jennifer Kizielewicz, Stavroula Kolitsopoulos, Sarah Kramer, Tatum Kutzer, Andria Kyriakides, Michael Landis, Abby Leap, Eugene Lee, Joann Lee, Hao Hong Lin, Stephanie Lloyd, Nikiesha London, Sofia Lopez, Shelby Lovechio, Michael Makarius, Tom Mandulak, Christopher Martino, Yvonne Martin, Michael McCarthy, Ruth McKee, Molly McIver, Shin Mitsugi, Yuliya Motuz, Christin Muller, June Nakajima, Ndeye Ndiaye, Laina Niciu, Shunsuke Nozawa, Helen Papaioannou, Joo-young Park, Eleni Passias, Benjamin Perriello, Jackie Pinkas, Karen Polanco, Amy Posner, Christopher Potts, Chris Poultney, Elizabeth Pratt, Daniela Raik, Peter Raucci, Alex Raytburg, Tiffany Robertson, Eleni Sakellis, Jesse Salazar, Christina Schlegel, Lina Shah, Linda Shipley, Sara Simonetti, Juyoun Song, Brian Stillman, Brian Struck, Natalie Toussaint, Suzanna Touzet, Adrienne Tremblay, Tony Tsai, Polina Veksler, Kathryn Venverloh, Jon Villanueva, Polina Veksler, Masaki Yamataga, and Terra York. I am grateful to them and to Maryam Bakht-Rofheart, Richard Cameron, David Heap, Erez Levon, and Linda Susman for helpful comments.

illustrate the fact that the quotative use of *go*, *like*, and *all* has become routine in the speech of many Americans:

- (1) He goes, "How are you doing in school?"
- (2) And I'm like, "I don't know."
- (3) He's all, "No!"

The change to new forms, particularly to *(be) like*, has been so pervasive and so swift as to prompt a number of studies, including Butters 1982, Blyth et al. 1990, Romaine & Lange 1991, Ferrara & Bell 1995, Dougherty & Strassel 1998, Sanchez & Charity 1999, Igoe et al. 1999, and Dailey-O'Cain 2000. (In contrast to these overall attestations of the strength of the spread is Cukor-Avila 2001, which reports a relatively low rate of quotative *like* use among African-American teenagers in a rural East Texas community.) Tagliamonte & Hudson (1999) study the presence of *(be) like* in Canada and the UK, while Macaulay (2001) focuses on Glasgow. Further, the calqued quotative *comme* has emerged in Montreal French (Sankoff 1993). With regard to American English, the present study's quantitative data constitute possibly the strongest evidence to date of the extent to which the new quotatives have taken over in the speech of those under 35, and especially those under 25.

This study's focus is quotative use in the New York City area, though the speakers are not all necessarily New Yorkers. Its focus is on *like*, *go*, and *all*. (As will be seen, the use of *all* continues to be highly infrequent in New York City and throughout the Northeast.) The data come from sociolinguistic interviews carried out by undergraduates in an introductory sociolinguistics course. As part of the course, I involve students in a class project, one that begins with each student recording two sociolinguistic interviews. The next stage involves the identification and coding of a particular sociolinguistic variable, preferably one that involves a change in progress. The students then extract the tokens from their interviews and code them for a set of factor groups. A graduate assistant and I check the coding and apply VARBRUL to the data. I then bring the results to class, the students and I discuss the findings and their implications, and students write up reports that assess both the findings and the project.

Quotatives were the focus of the class project in 1994 and again in 1995, 1996, 1997, and 1999. (I did not teach the class in 1998.) The findings of the class projects of 1995 through 1999 form the basis of the present paper, and all the examples and quantitative data in the present study come

from these corpora.<sup>2</sup> In each year, in those cases where a student's recordings yielded fewer than fifty quotatives, the student had the choice of carrying out an additional interview or of recording an unscripted program from television (most often, "trash" television). I have attempted to remove all of the television tokens and estimate that they now comprise less than one percent of the tokens in the present data base. The speakers under consideration are all native speakers of American English.<sup>3</sup> While there was no geographic limitation as to where in the United States the speakers came from, practicalities involving the speakers whom undergraduates might interview resulted in a situation in which the youngest speakers (aged 9 to 15) and those in the older groups (36 to 42 and, especially, 45 to 51) tended to be members of the interviewer's family and these, in turn, tended to live in the New York area. Thus, a greater percentage of the speakers under 15 and over 36 tended to come from the New York City dialect area than was true for speakers whose ages fell in between. Because speakers over the age of 51 use the new quotatives so rarely, particularly *like* and *all*, such individuals were not included in the study.

## 2 Why You Can't Do a VARBRUL Study of Quotatives . . .

The application of VARBRUL to quotative data apparently begins with Blyth et al. (1990). It is a general principle in the use of VARBRUL that, for a given set of variants, only those tokens should be included for which all of the variants are permissible. If a given environment blocks full variation, then that environment is excluded from tabulation. In variationist studies of the AAVE copula, for example, sentence-final tokens are routinely excluded from consideration because only full forms of the copula can occur in that environment (cf. Blake 1997). The distribution of quotatives in the corpora and subsequent confirmation from native speakers make the point

---

<sup>2</sup>The 1994 study was something of a pilot study. To the extent that the design of that study was comparable to that of subsequent years, it yielded very similar results.

<sup>3</sup>A "native" speaker of English was defined as one whose acquisition of American English began at age seven or earlier. The class project included fluent non-native speakers of American English as well, with a sharp distinction emerging between the quotative behavior of native and non-native speakers. While native speakers used *like* far more than any other quotative, non-native speakers used *say* most of the time. This is not to say that non-native speakers never used *like*: they used it 24% of the time (152/623). However, given the extent of the difference between native and non-native speakers and given the present paper's focus on the change that is occurring in native American English, the corpora in use in this paper are limited to data from native speakers of American English.

that there are several environments where quotative choice is constrained. Crucially, when that which is "quoted" is not speech itself, the choice of quotative is restricted. Simply put, a printed source *says*, e.g.:

- (4) The second page is a letter saying, "We regret to inform you . . ."  
 (5) So some of the people made up shirts that said, you know, "I saved Jack's butt."

a song *goes*, e.g.<sup>4</sup>:

- (6) How does that song go? "Those were two more friends of mine, they died."

and a gesture, a facial expression, or non-speech sounds *are all, are like, or go* but do not *say*, e.g.:

- (7) He didn't know how to use a fan and he was like, "[moves head]."  
 (8) We'd look at each other and go, "[makes face]."  
 (9) And he fills it up with water, and he goes, "[makes a blowing noise]."  
 (10) She dropped her book and she made this really weird noise, she's like, "Ablaaaaaaceehh!!"

In these cases where the choice of quotative is constrained by the type of material being quoted, it may be appropriate to distinguish between categorically non-occurring and strictly ungrammatical. That is, while quotatives other than *go* when the source is a song and quotatives other than *say* when the source is the printed word do not show up in the corpora (with one exception), to use some other quotative with songs or printed material does not seem to be patently ungrammatical in the way that the use of *say* with a gesture, facial expression, or speech sound is.<sup>5</sup>

Apart from differences in the distribution of quotatives when the quoted material is not speech, there are certain grammatical constructions that re-

---

<sup>4</sup>When referring to a song, the melody and the lyrics alike *go*. That is, the question, "How does that song go?" can be answered by humming, by speaking the lyrics, or by singing the lyrics. In some cases in the present corpora, *go* is immediately followed by *like*, e.g. *The song goes like, "Now that we found love."*

<sup>5</sup>From a quantitative perspective, I consider fewer than 5% to constitute "categorical non-occurrence." As indicated, there is one instance in the corpora (out of 22 total) where the printed word is quoted and for which *go*, not *say*, is used: *I like the one that goes, "Surgeon General's warning: Smoking can cause emphysema, heart disease..."*

quire specific quotatives. For example, if the subject of the quotative is a dummy *it*, then *like* is required:<sup>6</sup>

- (11) Sometimes it's just like, "God, what is this?"  
 (12) It's like, 'Let's go to the town board meeting; it's more exciting than Monday night wrestling,' Robert Legacy, a resident, said at a recent town board meeting. (*New York Times* 9/18/00)  
 (13) "I loved when he [Mark Jackson] went at Childs in the paper," Van Gundy said. "When he went at him, it was like, 'I got pride.'" (*New York Times* 3/19/01)

A second syntactic environment where only *like* can occur involves those instances where the quotative is not part of a verbal construction, i.e. is neither a verb itself nor teamed with a copula. To be sure, some AAVE speakers use *say* as a complementizer, but only when the verb of the higher sentence is itself a *verbum dicendi*, most often *tell*, e.g.:

- (14) They tell him say, "You better not go there." (Martin & Wolfram 1998:15)

However, *say* as a complementizer seems to be limited to this environment and, among American dialects of English, limited to AAVE. In contrast, *like* occurs with a wide range of higher verbs, as the examples from the present corpora illustrate:

- (15) And then I went through all this guilt like, "Oh, I must be such a bad person."  
 (16) I was pretending I was a reporter like, "Oh, we are here to. . ."  
 (17) I was running around the office like, "Oh my God, there's a cop on the line!"  
 (18) I was like the quasi-supportive friend like, "Oh, it really doesn't look that bad. No one'll notice. Don't worry."  
 (19) He was always the one like, "I'm upset."

It is also possible for quotative *like* to appear without a higher sentence:

- (20) Remember Jane? Fucking no-style, like smelled-weird Jane? Like, "Oh my god! You cut your bangs! They look great, Jane! Great, Jane, great!"

---

<sup>6</sup>Arguably *it's like* evolved from *it's as if X said*, where X ordinarily has an indefinite referent (as in (11) and (12)) but can have a specific referent (as in (13)). The construction may have played a crucial role in *like*'s grammaticalization as a quotative.

Given the range of sites where *like* can appear and its range of functions (cf. Underhill 1988), particularly in the speech of the young, it might be argued that the *like* in examples (15) through (20) is not always a quotative marker. While there are enough examples in the corpora to suggest that this *like* is indeed quotative and, further, that it is developing complementizer status (cf. Romaine & Lange 1991:260-1), what is directly relevant for a consideration of VARBRUL is that, of the putative quotatives, only *like* can occur in this environment.

The types of token discussed thus far are summarized in Table 1.

	<i>like</i>	<i>all</i>	<i>go</i>	<i>say</i>
The quoted material has a printed source	No	No	No	OK
The quoted material is a song	No	No	OK	No
What is quoted involves gesture, facial expression, or non-speech sounds	OK	OK	OK	No
The quotative's subject is dummy <i>it</i>	OK	No	No	No
The quotative used as complementizer	OK	No	No	No

Table 1. The grammaticality of particular quotatives in selected situations

Each of the token types in Table 1 represents an environment where not all quotatives occur. Tokens of these types are readily recognizable, can therefore be removed from a corpus, and thus do not threaten the validity of a VARBRUL analysis. Accordingly, the corpora under consideration in the remainder of this paper are ones from which such tokens have been removed.<sup>7</sup>

Other types of token present a problem, however. Take, for example, the use of one of these quotatives to indicate a paraphrase rather than a true, "literal" quotation. In the case of *say*, the morphosyntax of an indirect quote is different from that for a direct one. As Schoroup (1982) observes, *go* can only be used with direct quotes. In the case of *like* and *all*, a paraphrase looks just like a direct quote. The analyst (like the listener at the moment of speech) often has no way of knowing whether or not it was the speaker's intent to present verbatim speech. To be sure, even 'verbatim' quotes may represent some departure from what was actually uttered in the first place. This is why Tannen (1986) refers to them as "constructed dialogue." Still, with *say* and *go*, there is an attempt to approximate literalness. With *like* and *all*, on the other hand, no such attempt is required.

Further, a feature of *like* and *all* (discussed in Blyth et al. 1990:215 and elsewhere) is their ability to represent "inner monologues," i.e. what someone thought at a particular moment rather than what someone actually said.

<sup>7</sup>Of the tokens that were removed because they fell within one of the categories in Table 1, 167 were *like*, 25 were *go*, 21 were *say*, and 6 were *all*.



This is not possible with *say* or *go*. Because *like* and *all* can be used to report uttered and unuttered statements alike, one often cannot tell what the status of a given "quoted" statement is. Did someone actually say something or merely think it? For the examples in (21)-(29), although context may suggest that they were not actually said, one can't be sure.

- (21) He walked in and I was like, "Oh no, I am not seeing this!"  
 (22) I look at my knee and I was like, "Oh shit! It's the size of my head!"  
 (23) I'm like, "What is his problem?"  
 (24) I was so disgusted when the last one was stolen that I was like, "I'm not going to put another one in."  
 (25) I thought she was gonna be nice and stuff, so I was like, "Oh, Noriko's mean, so I'll just compensate."  
 (26) Then they all graduated, so I was like, "Oh what am I gonna do?"  
 (27) The guy just stared at me and goes, "I can fix that for you."  
 I couldn't believe he was like, "Oh, I'm Mr.Fix-it Man."  
 (28) She's like, "I'm it. Look at me. I shine."  
 (29) And she's like, "Oh my God!"

Note further that the examples in (27) through (29) do not involve first-person subjects. Still, because narrators are most likely and best able to report their own inner monologues, the "inner-monologue" cases are most likely in first person. As a consequence, one expects in turn that quotatives that have as one of their functions the expression of inner monologues will show up more frequently with first-person subjects than elsewhere. While the numbers from the NYU corpora in Table 2 confirm this, the difference in frequency between first person and second/third person is far smaller than in the earliest studies of *like*.

	1st person		2nd and 3rd person	
	n	%	n	%
1995	443/ 685	65%	576/1155	50%
1996	399/ 620	64%	395/ 842	47%
1997	339/ 602	56%	382/ 817	47%
1999	315/ 493	64%	383/ 681	56%
Total	1496/2400	62%	1736/3495	50%

Table 2. Distribution of *like/all* by the quotative's subject

As Ferrara & Bell suggest (1995), this appears to reflect the extent to which the grammaticalization of *like* has proceeded, specifically its emergence as an

all-around quotative rather than merely or primarily a quotative available for expressing inner monologues.<sup>8</sup>

In theory, a VARBRUL study ought to measure the likelihood that a given variant will obtain in a given environment. In a study of the overt marking of direct speech, one could be sure of which tokens of *say* and *go* to count but one could not be equally sure of which of tokens of *like* and *all* to count; this is the case because it is often not possible to know whether given instantiations of *like* and *all* signal verbatim speech, a paraphrase of verbatim speech, or unuttered thoughts. This problem for the quantitative analyst points to a central fact about the quotatives, namely that the change in progress is not merely a replacement of some forms by other forms.

For the listener, establishing whether the quoted material has been represented literally, has been paraphrased, or in fact constitutes an unuttered thought is usually not important. (When it does matter, the listener can question the speaker to find out.) In all three cases, i.e. literal and paraphrased and unuttered, the use of *like* or *all* functions to convey the speaker's attitude or emotions or perspective at the moment. Whether or not its use simultaneously conveys a speaker's utterance—and conveys it verbatim—is secondary. With *like* and *all*, the spirit of what is reported has become more important than the letter. That is the gist of the change in progress. Phrased another way, different quotatives make different claims. *Say* and *go* claim "literal" speech, while *like* and *all* do not. Indeed, when speakers use *like* or *all*, they are not even claiming that the speech in question ever actually occurred.

The reason that VARBRUL is not appropriate for the study of quotatives is that different quotatives make different claims. *Say* and *go*, on the one hand, and *like* and *all*, on the other, are not equivalent. For that reason, they are not truly variants of one another. *We are happy* and *we're happy* and *we happy* all are equivalent in meaning in AAVE, but *He walked in and I said, "Oh no, I am not seeing this!"* may well have a different meaning and different consequences from *He walked in and I was like, "Oh no, I am not seeing this!"* With *said* in this case, one can reasonably ask the speaker, "And what did he say when you said that?" With *like*, that's a highly unlikely, possibly infelicitous response. In sum, the reason that, strictly speaking, VARBRUL is not appropriate for the study of quotatives is that the change in quotatives has not simply been a change as to which quotative gets used. Rather, a more fundamental change has occurred, a change in the domain of usage. If by "variants" we mean alternative forms that are equiva-

---

<sup>8</sup>Ferrara and Bell also hypothesize that, early on, *like* was used primarily for dramatic effect and to signal internal dialogue; its use with direct speech represented a later step. It will be seen that, by that reckoning, *like*'s integration into the grammars of the speakers in the NYU corpora is quite advanced.

lent in meaning, then *say/go*, on the one hand, and *like/all*, on the other, are not congruent in domain, hence are not variants of each other.<sup>9</sup>

### 3 . . . And What Such a Study Can Show Us

Though the nature of quotatives in American English precludes a tightly constructed VARBRUL-based examination of their use, a quantitative study of quotative use is still revealing. In many but not all instances, the distribution of quotatives is so sharply delineated that the sophistication of a statistical program is unneeded, with raw frequencies alone being sufficient. In others, VARBRUL is appropriate. In every case in what follows, I have tried to reckon with the impact of the fundamental problem with the statistics, i.e. the unequal domains of the quotatives.

#### 3.1 *All and like (and all like)*

In the comparison of *like* and *all*, however, there does not appear to be any statistical problem. Their domains are identical, with each able to occur not only in reporting verbatim speech but also in providing paraphrases and reporting inner monologues. The NYU corpora show that New York is *like* territory, with *all* and also *all like* barely present. In contrast to the 3,233 occurrence of *like* (55% of the 5,898 tokens in the four corpora combined), there are only 38 occurrences of *all* and 27 of *all like*. The status of *all like* isn't fully clear; possibly it is a transitional form, part of a shift from *like* to *all*. The pair in (30)-(31) illustrate *all like*'s use:

(30) She was all like, "Oh, Hari, I miss you so much."

(31) I was all like, back with her, "I miss you too."

To the extent that *all* shows up in the corpora at all, it appears primarily in the speech of female college students, particularly Asian-Americans. While *all like* is also primarily used by college-age individuals, there are no particular patterns to its use with regard to sex or ethnicity.

Macaulay notes that all of the *all* users in Igoe et al.'s study at the University of Pennsylvania were in fact Californians. He adds, "There is no evidence so far that this form has been adopted further east" (2001:6). In the NYU corpora, most of the Asian-American college students who use *all* are from New York City and its environs. More than is true for college students generally, interaction among Asian-American students on college campuses

<sup>9</sup>VARBRUL is fundamentally a comparison of numerators when each numerator has the same denominator; with quotatives, the denominator diverges radically from numerator to numerator.

seems to be decidedly bicoastal. Thus, among upper-middle-class Asian-American college students at least, if *all* is going to expand beyond California, the Northeast is a likely next site.

In their University of Pennsylvania corpus, Igoe et al. (1999) found that *all* occurred with greatest frequency with first-person subjects. They linked that finding to *all*'s ability to present inner monologues. The NYU data do not support this finding at all. Rather, third-person subjects predominate, accounting for 76% of the *all* tokens and 78% of the *all like* tokens. *He's all* and *she's all* are especially likely; present-tense tokens with one of these two pronouns account for more than half of the occurrences of *all*. I do not have an explanation as to why third-person subjects generally and *he/she* specifically favor the selection of *all*. However, it is worth noting the parallel between *he's/she's all* and *he/she goes*. As I discuss subsequently, quotative *go* is by far most common in precisely those forms.<sup>10</sup>

### 3.2 Quotative Choice

To give the reader the full scope of the distribution of quotatives in the NYU corpora, I present the probabilities for *like* in Table 3 and the probabilities for *go* in Table 4. I give the full set of quotative frequencies by factor group in an appendix.

#### Speaker's Age

9-15	.62
18-24	.59
27-33	.52
36-42	.11
45-51	.12

#### Most Recent Quotative

<i>like</i>	.68	<i>go</i>	.35
first quotative in turn	.45	<i>say, other, all</i>	.28

#### Subject

1st person	.60	2nd and 3rd	.43
------------	-----	-------------	-----

#### Interviewer's Sex/Speaker's Sex

female/female	.58	male/female	.44
female/male	.46	male/male	.37

#### Tense

present	.57	past	.45
---------	-----	------	-----

<sup>10</sup>Throughout the remainder of the paper, quantitative references to *all* include both the 38 occurrences of *all* and the 27 occurrences of *all like*. Similarly, the nine occurrences of *go like* (e.g. *My friend Sonia goes like, "Say, 'God bless you'"*) have been combined with the 445 tokens of *go*.

**Speaker's Ethnicity**

Asian-American	.57	African-American	.40
other American	.50		
input	.57, $p = .000$		

Factor groups are presented in decreasing order of statistical significance.  
Table 3. Probabilities for (*be*) *like* for the four NYU corpora

<b>Tense</b>			
present	.79	past	.25
<b>Most Recent Quotative</b>			
<i>go</i>	.85	first quotative in turn	.45
<i>say</i> , other	.54	<i>like</i>	.45
<b>Subject</b>			
<i>he/she</i>	.64	1st	.36
other 3rd	.54		
<b>Speaker's Age</b>			
36-42	.62		
45-51	.53		
18-24	.51		
9-15	.48		
27-33	.41		
<b>Speaker's Sex</b>			
male	.55	female	.47
input	.04; $p < .006$		

Factor groups are presented in decreasing order of statistical significance.  
Table 4. Probabilities for *go* for the four NYU corpora

**3.2.1 Quotative Choice and Speaker's Age**

The most striking aspect of the usage of quotatives in the NYU corpora is the strength of the link between speaker's age and quotative choice. Young speakers use the new quotatives, predominantly *like*, most of the time, while speakers over 35 use them with relative infrequency. Certainly it is likely that speakers in the younger groups have, compared to those in the older groups, an inflated number of quotatives. That is, the younger groups' verbatim quotations, their paraphrases, and their inner monologues all are counted when they are introduced by *like*. In contrast, because speakers in the older groups don't use a quotative to present paraphrases and don't use one often for inner monologues, those speakers show up with appreciably fewer tokens overall. This difference in speakers' strategies has direct and meaningful consequences for tabulation. With that point acknowledged, consider the difference between speakers aged 9-15 and 45-51. The youngest group uses *like* 69% of the time, while the oldest uses it only 13%. (It is this figure for

the oldest group, 13%, that is unexpectedly high. Below I discuss the source of this result.) Within the 9-15 group, the use of new quotatives shows up in speaker after speaker. Thus, among the seven speakers in this group in the 1995 corpus who have 25 or more tokens, every speaker uses new quotatives, i.e. *like* or *all* or *go*, at least 62% of the time, as Table 5 shows.

Speaker's sex	<i>like</i>	<i>all</i>	<i>go</i>	Total %	N
female	100%			100%	78
male	95%			95%	63
female	74%	2%	19%	94%	53
female	91%			91%	33
female	81%	3%	5%	89%	37
female	46%	4%	27%	77%	26
female	62%			62%	29

Table 5. New quotative distribution by speakers 9-15 in 1995 corpus (n>25)

Each corpus was constructed such that all speakers came from one of five age groups: 9-15, 18-24, 27-33, 36-42, and 45-51. The division by age group was arbitrary. Those older than 51 were excluded because such speakers rarely use the new quotatives, especially *like* and *all*. The age groups rolled; that is, students in 1995 interviewed individuals who fit into those age groups in 1995, while students in 1999 interviewed individuals who fit into those age groups in 1999. The figures for quotative distribution by speaker's age across the four corpora make the point that the change in quotatives represents a generational change par excellence.<sup>11</sup> Beginning in 1999, an added category was created for bald quotatives, i.e. instances where there is no introduction of the quoted material, not even *I'm* or *she's*. In these cases the sole indicators of quotation are prosodic. In the 1999 corpus 8% of the tokens are bald. Vast individual variation obtains, with most speakers using bald quotatives rarely or not at all, while a few use them frequently. Out of 36 speakers in the 1999 corpus, four speakers provided more than half of all the bald tokens. Though more data need to be gathered in this regard and while bald-quotative use seems to be a matter of individual style, the data appear to support Sanchez & Charity's (1999) observation that frequency of bald-quotative use is greater among older speakers. Because the other corpora had not explicitly included bald tokens, some were included but probably a number of others were omitted (since this was, after all, a study of quotatives and bald quotations lack any overt quotative). Accordingly, in the

<sup>11</sup>A given student's two interviews came from speakers from two different age groups. There was a sign-up sheet in order to ensure that all the age categories were filled. A larger number of interviews was permitted with younger speakers.

present study, all recognizably bald tokens have been removed from consideration.

Frequencies and probabilities alike show there to be a dramatic drop-off in *like* use between the 27-33 and the 36-42 groups. The 36-42 and 45-51 groups each have one "young" speaker, i.e. an individual whose quotative use displays the pattern characteristic of younger speakers. It is these two speakers who make the new quotative numbers as high as they are for the two older groups. Table 6 illustrates this point.

Speaker	Age	<i>like</i>	<i>all</i>	<i>go</i>	total %	n
the "young" one	36-42	57%		2%	60%	42
the "young" one	45-51	56%	2%	20%	77%	84
18 others	36-42	6%		12%	18%	394
23 others	45-51	2%		5%	7%	342

Table 6. Quotative distribution among speakers in the two older age groups, all four corpora combined.

A critical part of constructing one's identity involves age. That is, a range of social patterns and norms exist that define and delimit age-appropriate behavior, including age-appropriate linguistic behavior. Both of the speakers in Table 6 who differ from their agemates in their use of quotatives can be said, I would argue, to be departing from and possibly violating social norms thereby. The "young" speaker in the 36-42 group is possibly only a few years older than those for whom *like* is the usual quotative; consequently, his frequent use of *like* may not be particularly noticeable. On the other hand, the "young" speaker in the oldest age group is dramatically different from her agemates. She is an elementary school teacher, someone who interacts not only with young students but also with teachers who are much younger than she is. These biographical facts explain her exposure to the new quotatives but not necessarily her decision to use them. While a majority of her quotatives are *like* and she even uses *all like* twice, none of the sixteen other speakers in her age group with ten or more quotative tokens have more than two occurrences of *like*; most of the sixteen have no *like* tokens at all.<sup>12</sup>

The simplest view of the generational change in progress is that it is both straightforward and inexorable, with the prevalence of *like* growing as individual users of *like* grow older and the percentage of primary *like* users in the general population increases thereby. By that view, the sharp drop that separated the 27-33 group in the corpora from the 36-42 group represents the

<sup>12</sup>When the two "young" speakers in the older groups are removed, the probabilities for *like* for speaker's-age factor group become the following: 9-15, .64; 18-24, .61; 27-33, .54; 36-42, .07; 45-51, .04.

temporal "isogloss" of the change. This reasoning has it that in ten years the sharp drop will separate a 37-43 group from a 46-52 group.

In fact, it may not be as simple as that. While this is certainly a generational change in progress, there may well be more going on. The present link between quotative use and a speaker's age carries with it a prescriptive disapproval of the "young" quotatives. If using the new quotatives expresses the speaker's youth, a switch to more conservative quotatives may signal a more "adult" approach to the world. (The assignment of the "young speaker" label to the person who uses *like* as a quotative may be intensified by the popular assignment of *like* the filler and *like* the focus marker (cf. Underhill 1988) to the speech of young individuals.) In other words, age grading may be present. In 1982 Butters observed that *like* use was largely restricted to those under thirty. Now, almost twenty years later, heavy *like* use continues to be restricted to those under 35. Still, it remains to be established definitively that individuals curtail their use of *like* at some point when they are in their thirties. Further, if indeed this is the case now, it is not yet clear whether this is a temporary phenomenon, one that retards but does not block the change in progress, or one that will become a more permanent feature of the social dimensions of *like* usage. Even if there is an element of age grading at work, *like* and *all*, by virtue of their use with paraphrases and inner monologues, have reconfigured the domain of quotative usage. This fact makes it likely that they have entered—or will soon enter—a fixed place in the vernacular. For contrast, consider post-clausal negation in English with *not* (or with *psych*), as in (32)-(33), a phenomenon that has proven to be a syntax fad.

(32) Yeah, all of us can fit into this elevator, not!

(33) I wanna be just like you, not!

Even though post-clausal negation is found in other languages, e.g., Ewe, and despite its popularity among American youth in the late 1980's and early 1990's, it failed to enter the language in a lasting way. Its use in American English was so highly marked (with its intent to fool the listener, at least for a moment) that it never became grammaticalized. Further, apart from its element of momentary deception, it was negation as usual. As such, it was easy for its usage to wane once the novelty had worn off. On the other hand, *like* and *all*—because of the way in which they have shifted focus from the letter of quotations to their spirit—are likely to become a permanent part of the vernacular if they have not already done so.

In any event, within the NYU corpora, the focal point of the change in progress would seem to be the middle group, i.e. the speakers aged from 27 to 33. When the two "young" speakers of Table 6 are removed from consideration, the speakers in the 27-33 group show the greatest range *internal* to



an age group. The speakers listed in Table 7 show just how extreme the range is.

Speaker's sex	Corpus	<i>like</i>	<i>all</i>	<i>go</i>	total %	n
female	1997	100%			100%	29
female	1995	96%	3%	1%	100%	98
female	1999	95%		2%	97%	43
female	1996	30%		5%	35%	64
female	1995	16%		14%	30%	37

Table 7. Quotative distribution in the four corpora: speakers in the 27-33 group with the greatest use or non-use of the new quotatives (n>25)

The general orderliness of *like* with respect to speaker's age does not obtain for *go*. For speakers 36 and over (always with the exception of the two "young" speakers), *go* continues to cling to its place as the primary new quotative. For speakers 33 and under, this has ceased to be the case.

### 3.2.2 Quotative Choice and Linguistic Factors

Apart from the quotative's subject pronoun, two other linguistic factor groups were tested and are also statistically significant, preceding quotative and verb tense.<sup>13</sup>

As its name suggests, the "previous quotative" factor group refers to the immediately preceding quotative, provided that it occurred in the same turn. The results suggest that there tend to be strings of the same quotative. Verb tense proves to be statistically significant as well, but the relationship between tense and quotative choice is most likely not causative. Rather, present tense and the new quotatives co-occur because both reflect greater informality. That is, occurrences of the present tense in the corpora are almost always instances of the *historical* present, a tense that is characteristically informal and colloquial.<sup>14</sup> Similarly, *like* as a quotative is informal and colloquial.<sup>15</sup> The same explanation extends to the correlation between present tense and quotative *go*. A further point with regard to *go* is the great frequency of *he goes* and *she goes*: of all the present-tense tokens in the corpora

<sup>13</sup>"Previous quotative" is coded only for the 1995, 1996, and 1999 corpora.

<sup>14</sup>"Present" tense is more accurately non-past. This classification included both basic non-past and also present progressive. "Past" tense included past preterit, past habitual (whether *used to* or *would*), and past progressive.

<sup>15</sup>In the case of *go*, it seemed desirable to separate second-person subjects from third-person ones; however, there were too few second-person tokens for it to test reliably (166/5895, 3%). As a result, in this instance the second-person subject factor has been removed.

with *he* or *she* as the subject, fully 22% are *he goes* or *she goes*. While I don't find this result surprising, I have no explanation to offer for it.

### 3.2.2 Quotative Choice and Social Factors: Sex and Ethnicity

The quotative literature varies as to correlations between speaker's sex and quotative use. Blyth et al. assert that men are more likely to use *like* quotatives, while Romaine and Lange, Igoe et al., and other studies get the opposite result. The NYU corpora show that males favor the quotative *go* (.55) while females disfavor it (.47). In the case of *like*, speaker's sex provided less a fit than did the sex of the two participants in the conversation, i.e. both speaker and interviewer. Same-sex dyads show strong results, with female pairs favoring *like*, male pairs strongly disfavoring it, and mixed-sex pairs weakly disfavoring it.

Speakers' ethnicity was also tabulated, with most speakers identified as African-American, Asian-American, or "other", with the latter almost always European-American. ("Asian-American" refers to those of Chinese, Korean, or Japanese heritage.) While ethnicity was not statistically significant for *go*, it was for *like*, with Asian-Americans favoring the use of *like* (.57) and African-Americans disfavoring its use (.41). Within the corpora there are no Asian-American speakers in either of the two older age groups. This raised the possibility that the seeming correlation between Asian-American ethnicity and greater use of *like* had arisen from a statistical skew. To test for this possibility, I examined a subset of all the tokens, including only those that came from speakers in the three younger age groups. The results show no weakening of the distribution; rather, they show a very slight strengthening of it, with probabilities of .58 for Asian-Americans, .39 for African-Americans, and .50 for the "other" group. Further study is needed as to why Asian-Americans use *like* with greater frequency than other speakers. I noted earlier that Asian-American college students appear to be the first New Yorkers to be using quotative *all*. The two phenomena, greater Asian-American use of *like* and Asian-American leadership in the use of *all*, may arise from a single cause. If both quotatives originated in California (and it appears that they did) and if the New York-area Asian-American college students' ties to California Asian-American college students are especially strong, these facts lay the groundwork for an explanation as to why Asian-American college students show the greatest use of *like* and *all*. I hypothesize that the primary source of the introduction of *like* and *all* into the speech of other college students in the Northeast has likewise been California, with the difference between Asian-Americans and others in the Northeast in their rates of adaptation of the California features reflecting a difference in the strength of their ties to California.

The fact that the African-American speakers in the corpora tend to disfavor the use of *like* might seem to go against, for example, Sanchez & Charity 1999, who demonstrate that *like* is vigorously present among urban African-American speakers.<sup>16</sup> In fact, *like* use among the youngest African-American speakers (9-15) in the NYU corpora is robust, with a frequency of 67% (compared to 69% for all other speakers in the age group). In the intersection of age and ethnicity, it is only among the 18-24 and 27-33 groups that African-American speakers show lower rates, 44% and 46% respectively versus 61% and 53% for other speakers. Sanchez & Charity carried out their research inside a community, while the African-American speakers in the NYU corpora are present as the result of individual, unrelated sociolinguistic interviews. Because an African-American speaker's vernacular, especially if it is AAVE, is subject to ready stigmatization, African-Americans interviewed for undergraduate sociolinguistic courses, *ceteris paribus*, may be less likely to use their vernacular than are other speakers.<sup>17</sup>

On the basis of a qualitative assessment, Butters (1989:149) argues with reference to *go* and *like* that ethnicity is not salient in determining quotative use. In that view, when it comes to quotatives, age matters, race doesn't. Regardless of whether the NYU corpora underestimate NYC-area African-American rates of *like* or reflect them accurately, the relevant fact that emerges from the NYU corpora and Sanchez & Charity's study alike is support for Butters's view: African-Americans are participants in the switch to the new quotatives.

### 3.3 Grammaticalization in Progress

#### 3.3.1 Evidence of Incomplete Grammaticalization

While I argue that the grammaticalization of the new quotatives is in progress, I would not claim that it has achieved its completion. At the 2000 NWAV, William Labov questioned the extent to which *like* can be used in questions and in negation. While *like* does occur in the speech of those 33 and under in the NYU corpora in both questions and negation, as in (34)

<sup>16</sup>The African-American speakers in Cukor-Avila (2001) show far lower rates of *like* use but, crucially, they are rural.

<sup>17</sup>While I think that issues of language and tape recording are especially relevant for the African-American speakers who consented to be recorded in the present case, I acknowledge that all the data in the NYU corpora come from in and around New York City, Labov's "great sink of negative prestige." Commenting on the self-evaluations that arise in the interviews that form the basis of *The Social Stratification of English in New York City*, Labov states, "The term 'linguistic self-hatred' is not too extreme . . ." (1982:344).

through (37), it does not do so often. (Regardless of the quotative, there are very few instances where the quotative itself was part of a question.)

(34) Were you like, "Tell me what you think of me"?

(35) I'm not like, "I'm going to go talk to her."

(36) We weren't like, "Let's take hours to make the store pretty."

(37) You couldn't laugh at it or just be like, "Get the fuck [out of here]."

I examined in more detail the distribution of negative quotatives among those speakers 33 and under in the 1999 corpus who had at least one negative quotative. Among these speakers, *like* accounted for 67% of all the affirmative quotatives (117/175) but only 29% of the negative ones (2/7). In contrast, *say* and other traditional quotatives provided only 30% of the affirmative quotatives but a full 71% of the negative ones.

### 3.3.2 Transitional Forms

The corpora contain evidence that suggests that the transition from *say* to *like* (with *go* serving as something of an intermediate stage) has involved a range of intermediate steps and experiments. For example, each of the corpora displays putative double quotatives. In 3.1 I discussed the possibility that *all like* might be intermediate between *like* and *all*. The double quotatives that I wish to consider now are possibly from an earlier evolutionary stage. In these cases, an existing quotative teams up with *like*:

(38) Before she was going out with W, she said like, "I would rather have a ball of my own pus than go out with W."

(39) Yeah, so I go like, "Don't change the subject."

(40) So I was telling K like, "You walk into the bank tomorrow and just tell them that you need to close my account for me."

This same pattern shows up with thought-verbs, e.g.:

(41) 'Cause I don't want people to think like, "What the hell? Is she up to her wrists in there?"

(42) Then he saw us standing on stage and saw me standing up on stage and then he kind of realized like, "Uh-oh. This is a conspiracy."

In some cases, the appropriate analysis of putative double quotatives is that they are not double at all. Rather the verb is the quotative, while *like* is functioning in a non-quotative way, e.g.:

(43) Well she just said, y'know, like, "What are you up to tonight?"

In others, however, there seems truly to be a double quotative, specifically a quotative verb and a quotative complementizer, as in some of the examples above and also the following:

- (44) Bill said like, "I don't even know why I took her out."  
 (45) She goes like, "You know something? You have herpes."

### 3.3.3 *I'm thinking*

In the course of studying quotatives, I became aware of my own use of *I'm thinking*. For those speakers who don't use *like* (ordinarily because these speakers are older than the habitual *like* users), this use of *thinking* provides an alternative way to present dialogue in the representation of inner monologues. *I thought* is already available, but *I'm thinking*—because it is in the historical present—is more informal, as these examples from speakers in the two older groups illustrate:

- (46) I'm thinking all cool, "Hey, look I can do this."  
 (47) Then I was also thinking, "Hey, you know . . ."

However, as Maryam Bakht-Rofheart (p.c.) points out, young *like* users make use of *I'm thinking* as well, as in the following examples:

- (48) I'm thinking, "She's young."  
 (49) I'm thinking, "I got this guy fired, or transferred."

A possible explanation as to why *like* users might use *I'm thinking* to introduce inner monologues is that *I'm thinking* is unambiguous in representing an inner monologue. The speaker thereby makes clear to the listener that the dialogue was not actually said.

## 4 Conclusion

In this paper, I have presented a set of reasons as to why the study of variation in quotative use does not lend itself to VARBRUL-style number crunching. The fact that one cannot do the usual kind of VARBRUL study of quotatives arises directly from the nature of the change in progress. This is not simply a change in forms. It is a change in domain.

Showing the limitations of VARBRUL for assessing quotative use is not to say that a properly located quantitative VARBRUL study has no contribution to make to our understanding of quotatives. Strictly speaking, statistical programs are always tools, never analyses. That truth is simply more critical in the present case.

A change in quotative use and quotative domain is now in progress. What will the outcome be? Is the new quotative *go* already an old quotative? Will *go* continue to spread, start to decline, or do both concurrently in different parts of the speech community? And what about *all*? Does *like* stand a chance against its California rival? For now, in the Northeast at least, *like* is on top.<sup>18</sup>

As I have suggested in the course of this paper, the fact that *like* and *all* have altered the domain of quotative usage increases the likelihood that one, if not both, will survive. It remains to be seen whether or not *like* will become so fully grammaticalized as to be the unmarked quotative in all environments, even with negation and questions. More remote but still a possibility is that *like* will become a freestanding complementizer with few or no co-occurrence restrictions on its higher verb, as in the examples in (15)-(20). Such a role is a long way off, there being few occurrences of *like* with this function in the NYU corpora. In terms of what has already happened with *like*: The speakers at the top of Tables 5 and 7 represent a phenomenon that shows up more widely in the corpora. They are the total *like* users, the speakers who use *like* 95% of the time or more. Arguably, they represent the vanguard, their speech testifying to the force with which *like* has established itself as the primary quotative of much of America's vernacular English.

### Appendix: Quotative Frequencies in the NYU Corpora

Subject	<i>Like</i>	<i>all</i>	<i>go</i>	<i>say</i> + other	n
<i>I</i>	1394 62%	7 0%	85 4%	768 34%	2254
<i>you</i>	57 34%	3 2%	9 5%	97 58%	166
<i>he/she</i>	1177 51%	33 1%	286 12%	821 35%	2317
<i>we</i>	102 70%	5 3%	2 1%	37 25%	146
<i>they</i>	222 55%	9 2%	20 5%	153 38%	404
3sg noun	234 46%	6 1%	46 9%	222 44%	508
3pl noun	<u>46</u> 46%	<u>2</u> 2%	<u>6</u> 6%	<u>46</u> 46%	<u>100</u>
	3232 55%	65 1%	454 8%	2144 36%	5895
	<i>Like</i>	<i>all</i>	<i>go</i>	<i>say</i> + other	n

<sup>18</sup>Of all the *like* tokens in the corpora, 6% of them are *just like*, e.g. *She was just like, "You didn't say what you really mean."* Further research is needed to address the range of questions that emerge: Is *just like* best understood at face value as the combination of *just* 'only' plus quotative *like*, or is it most often simply a variant of *like*? What, if any, is the difference in distribution between *like* and *just like*? Did *just like* have a distinctive role to play in the emergence of *like* as a quotative, and/or is it now serving some transitional function in *like*'s ongoing evolution?

<b>Tense</b>					
present	1547 62%	45 2%	382 15%	525 21%	2499
past	1570 51%	19 1%	41 1%	1430 47%	3060
other	<u>97</u> 43%	<u>1</u> 0%	<u>14</u> 6%	<u>113</u> 50%	<u>225</u>
	3214 56%	65 1%	437 8%	2068 36%	5784
<b>Most Recent Quotative</b>					
1st in turn	659 48%	19 1%	72 5%	626 45%	1376
like	1418 78%	19 1%	108 6%	270 15%	1815
go	113 45%	2 1%	85 34%	51 20%	251
say	213 27%	6 1%	62 8%	517 65%	798
other/all	<u>79</u> 43%	<u>2</u> 1%	<u>3</u> 2%	<u>101</u> 55%	<u>185</u>
	2482 56%	48 1%	330 7%	1565 35%	4425
<b>Interviewer/Speaker</b>					
fem/fem	1679 61%	24 1%	187 7%	870 32%	2760
mal/mal	371 43%	5 1%	74 9%	416 48%	866
fem/mal	654 53%	12 1%	128 10%	436 35%	1230
mal/fem	<u>529</u> 51%	<u>24</u> 2%	<u>65</u> 6%	<u>424</u> 41%	<u>1042</u>
	3233 55%	65 1%	454 8%	2146 36%	5898
<b>Speaker's Age</b>					
9-15	674 69%	3 0%	76 8%	230 23%	983
18-24	1990 63%	56 2%	254 8%	882 28%	3182
27-33	467 54%	4 0%	45 5%	355 41%	871
36-42	47 11%	0 0%	47 11%	342 78%	436
45-51	<u>55</u> 13%	<u>2</u> 0%	<u>32</u> 8%	<u>337</u> 79%	<u>426</u>
	3233 55%	65 1%	454 8%	2146 36%	5898
<b>Speaker's Ethnicity</b>					
Afr. Am	285 46%	1 0%	29 5%	308 49%	623
Asian Am	484 69%	16 2%	40 6%	161 23%	701
other Am	<u>2464</u> 54%	<u>48</u> 1%	<u>385</u> 8%	<u>1677</u> 37%	<u>4574</u>
	3233 55%	65 1%	454 8%	2146 36%	5898
<b>Excluded from VARBRUL: Tokens Occurring in Invariant Settings</b>					
	<u>167</u> 76%	<u>6</u> 3%	<u>25</u> 11%	<u>21</u> 10%	<u>219</u>
<b>Totals</b>	3400 56%	71 1%	479 8%	2167 35%	6117

## References

- Blake, Renée. 1997. Defining the envelope of linguistic variation: The case of 'don't count' forms in the copula analysis of African-American Vernacular English. *Language Variation and Change* 9:57-80.
- Blyth, Carl, Jr., Sigrid Recktenwald, and Jenny Wang. 1990. *I'm like*, "Say what?!": A new quotative in American oral narrative. *American Speech* 65:215-27.

- Butters, Ronald R. 1982. Editor's note [on *be like* "think"]. *American Speech* 57:149.
- Butters, Ronald R. 1989. *The death of Black English: Divergence and convergence in Black and White Vernaculars*. Bamberger Beiträge zur Englischen Sprachwissenschaft, 25. Frankfurt: Lang.
- Cukor-Avila, Patricia. 2001. "She say," "she go," "she be like": Verbs of quotation over time in African-American Vernacular English. Paper presented at the Annual Meeting of the American Dialect Society, Washington, DC.
- Dailey-O'Cain, Jennifer. 2000. The sociolinguistic distribution of and attitudes toward focuser *like* and quotative *like*. *Journal of Sociolinguistics* 4:60-80.
- Dougherty, Kevin A. and Stephanie M. Strassel. 1998. A new look at variation in and perception of American English quotatives. Paper presented at NWAWE 27, University of Georgia.
- Ferrara, Kathleen, and Barbara Bell. 1995. Sociolinguistic variation and discourse function of constructed dialogue introducers: The case of *be + like*. *American Speech* 70:265-90.
- Igoe, Matthew, Nel Lamb, Jonathon Gilman, and Ron Kim. 1999. Quotatives: The further grammaticalization of *be + like* and some observations on *be + all*. Paper presented at NWAWE 28, Toronto.
- Labov, William. 1982 [1966]. *The social stratification of English in New York City*. Washington: Center for Applied Linguistics.
- Macaulay, Ronald. 2001. *You're like "why not?"* The quotative expressions of Glasgow adolescents. *Journal of Sociolinguistics* 5:1-21.
- Martin, Stefan, and Walt Wolfram. 1998. The sentence in African-American vernacular English. *African-American English*, ed. Salikoko S. Mufwene, John R. Rickford, Guy Bailey, and John Baugh, 11-36. London: Routledge.
- Romaine, Suzanne, and Deborah Lange. 1991. The use of *like* as a marker of reported speech and thought. *American Speech* 66:227-78.
- Sanchez, Tara, and Anne Charity. 1999. The use of *be like* and other verbs of quotation in a predominantly African-American community. Paper presented at NWAWE 28, Toronto.
- Sankoff, Gillian. 1993. Anglophone bilingualism in Montreal. Lecture presented at NYU.
- Schourup, Lawrence. 1982. Quoting with *go* "say." *American Speech* 57:148-9.
- Tagliamonte, Sali, and Rachel Hudson. 1999. *Be like et al.* beyond America: The quotative system in British and Canadian youth. *Journal of Sociolinguistics* 3:147-72.
- Tannen, Deborah. 1986. Introducing constructed dialogue in Greek and American conversational and literary narrative. *Direct and indirect speech*, ed. Florian Coulmas, 311-32. Amsterdam: Gruyter.
- Underhill, Robert. 1988. *Like* is, like, focus. *American Speech* 63:234-46.

Department of Linguistics  
New York University  
719 Broadway, Rm. 501  
New York, NY 10003  
john.singler@nyu.edu