Available online at www.sciencedirect.com

ScienceDirect

Journal of Pragmatics 81 (2015) 1–7

journal of PRAGMATICS

www.elsevier.com/locate/pragma

Quotation as a multimodal construction☆

Natalia L. Blackwell [a], Marcus Perlman [b], Jean E. Fox Tree [a,*]

[a] University of California, Santa Cruz, USA
[b] University of Wisconsin, Madison, USA

Received 30 October 2014; received in revised form 26 February 2015; accepted 9 March 2015

Abstract

Quotations are a means to report a broad range of events in addition to speech, and often involve both vocal and bodily demonstration. The present study examined the use of quotation to report a variety of multisensory events (i.e., containing salient visible and audible elements) as participants watched and then described a set of video clips including human speech and animal vocalizations. We examined the relationship between demonstrations conveyed through the vocal versus bodily modality, comparing them across four common quotation devices (*be like*, *go*, *say*, and zero quotatives), as well as across direct and non-direct quotations and retellings. We found that direct quotations involved high levels of both vocal and bodily demonstration, while non-direct quotations involved lower levels in both these channels. In addition, there was a strong positive correlation between vocal and bodily demonstration for direct quotation. This result supports a Multimodal Hypothesis where information from the two channels arises from one central concept.

Keywords: Quotation; Quotatives; Demonstration; Gesture; Body; Prosody

Quotation is traditionally treated as a device used to report a person's speech (Coulmas, 1986). Yet this characterization does not cover the wide range of ways that quotation is often used to *demonstrate* various kinds of events, often involving visible qualities as well as audible ones (Clark and Gerrig, 1990; Fox Tree and Tomlinson, 2008; Hudson, 1985; Waksler, 2001). The multimodality of quotation is illustrated in the following examples of (1) a vocal demonstration of an audible event, (2) a bodily demonstration of a visible event, and (3) a combined vocal and bodily demonstration of a multisensory event:

(1) The car engine went [brmbrm], and we were off (Clark and Gerrig, 1990; from Hudson, 1985).
(2) I got out of the car, and I just [demonstration of turning around and bumping his head on an invisible telephone pole] (Clark and Gerrig, 1990; from Hudson, 1985).
(3) He was like opening his drawers but they were like stuck so he was like, ''Raah aah'' [growling sounds, while pretending to pull on a drawer] (Fox Tree and Tomlinson, 2008: 100).

☆ This research was supported by faculty research funds granted by the University of California Santa Cruz. We thank our many research assistants who aided in data collection and coding, with a special thanks to Annecy Majoros, Jordan Martin, Sarit Fassazadeh, Carly Johnson, and Aza Tetelman.

* Corresponding author at: Psychology Department, Social Sciences II, University of California, Santa Cruz, CA 95064, USA. Tel.: +1 8314595181.
E-mail addresses: nblackwe@ucsc.edu (N.L. Blackwell), mperlman@wisc.edu (M. Perlman), foxtree@ucsc.edu (J.E. Fox Tree).

http://dx.doi.org/10.1016/j.pragma.2015.03.004
0378-2166/

Here we examine how vocal and bodily demonstrations are coordinated in direct and non-direct quotations, including quotations that followed three common quotative devices, *be like*, *go*, and *say*. Our investigation sheds light on quotation as a multimodal construction, and more generally, on the nature of language as a multimodal system of communication.

Quotation. The canonical, *direct* form of quotation allows a speaker to convey another person's words such as in the following example (this and all further examples in this paper are from the current corpus):

(4) He said, ''After a while, crocodile.''

Quotation can also be used for other, non-speech sounds, such as the pre-linguistic babbling of an infant, the vocalization of a nonhuman animal, or the crack of an inanimate firecracker:

(5) The babies were going, ''Ga ga ga da da.''
(6) The cows were like, ''Mooo.''
(7) It goes, ''Pop pop pop pop pop pop pop.''

Although direct quotations are sometimes thought of as veridical reports of what was said, they are not necessarily verbatim (Blackwell and Fox Tree, 2012; Fox Tree and Tomlinson, 2008; Wade and Clark, 1993; and for discussion, see Clark and Gerrig, 1990; Romaine and Lange, 1991). Instead, they appear to be used to create imagined experiences (Pascual, 2014), or to recreate for the listener ''what it would be like to hear, see, or feel what the original speaker did'' (Wade and Clark, 1993: 818). Quotations are often associated with dramatic presentation and elevated levels of sensorimotor imagery. For example, speakers often produce direct quotations for dramatic effect at peak points in a narrative (Li, 1986), and in turn, listeners find narratives to be more dramatic and memorable when presented as direct, first-person speech (Schiffrin, 1981; Tannen, 1989). In addition, reading direct quotations, but not indirect quotations, activated areas of the brain related to voices (Yao et al., 2011). Actual reading behavior was also affected – reading rate was faster when quoting a fast-paced speaker as compared to a slow-paced speaker (Yao and Scheepers, 2011).

In English and other languages, direct quotations are frequently introduced by particular lexical devices, or *quotative*s, such as *tell*, *ask*, *say*, *be like*, and *go*. A number of factors contribute to the choice of a particular device, including the nature of the quoted content as well as social factors. Compared to *say*, the use of *be like* is associated with looser reports of the original quoted content and with the expression of attitude or emotion (Andersen, 1998, 2000; Blyth et al., 1990; Romaine and Lange, 1991), although both of these proposals have been challenged (Blackwell and Fox Tree, 2012; Fox Tree and Tomlinson, 2008). Compared to *be like*, the use of *say* is associated with higher status speakers being quoted and higher status addressees (Blackwell and Fox Tree, 2012). Both *be like* and *go* are associated with the reporting of non-speech sounds and also demonstrable actions and gestures (Butters, 1980; Buchstaller, 2001; Fox Tree and Tomlinson, 2008). In some cases, quotations are produced without any introductory device to frame the quoted material (i.e., *zero quotatives*). The lack of any verb requires speakers to use more vocal modulation to signify the act of quotation, and zero quotatives are often used when conveying a more dramatic representation of the quoted speech (Mathis and Yule, 1994).

Gesture and quotation. Previous researchers have shown a hand-in-hand relationship between the amount of information conveyed through the words of an utterance and the amount of information conveyed through co-speech gesture (de Ruiter et al., 2012; So et al., 2009). Utterances that are more informative by words also tend to be more informative by gesture. However, under certain circumstances there can also be a trade-off between speech and gesture. When speech is more effortful, such as when referring a partner to mutually visible targets that are more difficult to distinguish verbally, gesturing takes over some of the cognitive load. When gesturing is more effortful, as when the targets are farther away, speech takes over (Bangerter, 2004; de Ruiter, 2006; van der Sluis and Krahmer, 2007).

The production of quotations in particular is associated with a high rate of gesturing, and it is often coordinated with distinctive patterns of eye gaze and body orientation (Goodwin, 2007). The production of quotations is also associated with vocal demonstrations, so that words are quoted with additional vocal qualities, such as enacted lisps or drawls, and non-speech sounds are quoted as vocal imitations of the originals (Clark and Gerrig, 1990; Wade and Clark, 1993). Yet, when speakers use quotations to demonstrate an event, especially non-speech events, the question remains as to how they coordinate the vocal and bodily channels. That is, we know that multimodal communication can take place in a variety of ways, such as hand gestures and words indicating different aspects of the meaning of a concept (Kendon, 2014). We ask how vocal demonstration and bodily demonstration coordinate with each other.

A *Multimodal Hypothesis* predicts that vocal and bodily demonstrations arise from a single underlying representation of the quoted content, one that does not distinguish between sensory modality. Thus demonstration through the vocal and bodily channels would both rise concomitantly with a general rise in the overall degree of demonstration. Alternatively, according to a *Trade-Off Hypothesis*, a demonstration is produced preferentially through one channel or the other.

Quotation potentially poses an increased cognitive load on the speaker who must reproduce a relatively detailed account of the quoted event, often in the context of presenting a story. Thus the cognitive effort required to produce a vocal demonstration may detract from the level of bodily demonstration, and vice versa, such that the more one type of demonstration is used, the less the other is used. Finally, according to an *Independent Hypothesis*, there might be no relationship between these two modes of demonstration.

Current study. In the current study, we examined the use of quotation as a multimodal construction to report an expanded variety of speech and speech-like events, including accented human speech, pre-linguistic babbling of babies, speech-like vocalizations produced by animals, and speech produced by robots. Each of the events also contained salient visible elements. The main purpose was to investigate how speakers coordinate demonstration in the vocal and bodily modalities. This would determine which of the three hypotheses, *Multimodal*, *Trade-Off*, or *Independent*, best describes the way in which such kinds of events are quoted. In addition, we investigated whether there was a relationship between the level of demonstration and the selection of particular quotation devices.

1. Method

1.1. Participants

One hundred and thirty-seven undergraduates at the University of California, Santa Cruz, participated in the experiment in exchange for course credit.

1.2. Materials

Participants watched eighteen short YouTube videos (40 s–2 min 20 s in length) involving an assortment of quotable sounds and actions. The videos contained a variety of vocalizations by humans (adult and infant) and nonhuman animals, sounds produced by inanimate objects, and a variety of movements.

1.3. Procedure

Participants described each video to a confederate addressee directly after viewing it. They were instructed to be descriptive and entertaining. The confederate was allowed to laugh and provide appropriate back channels. The descriptions were video-recorded and transcribed for analysis. Individual descriptions were between 20 s and 4 min (M = 1.73 min, SD = 3.21 min).

1.4. Analysis

Analysis of the video-recordings proceeded in two steps. In the first step, stretches of recordings that met certain criteria were selected. In the second step, ratings of these stretches of recording (*clips*) were obtained.

Quotation selection. To eliminate variability from the quotation of different content within a video, we focused analysis on descriptions of a single quotable event for each clip (the primary audible event) and included only quotations related to that particular event. To eliminate video-quotative confounds, we further focused our analyses on the clips that elicited uses of each of three verbal quotatives: *be like*, *go*, and *say* (*just* was scarcely used and not evaluated further). If *goes* was always used for one video and *says* for another, we would not be able to tease apart the contribution to demonstration that resulted from the use of a particular quotative versus a particular video. Ten clips did not elicit instances of each of these quotatives, leaving eight clips for analysis: (1) a dog howling what sounded like the words *I love you*, (2) a parrot producing imitations of human speech and a tiger's roar, (3) another parrot counting to ten, (4) two infants who appeared to be talking to each other, (5) a talking robot, (6) a speaker pretending to have a French accent, (7) a woman falling during a news interview about grape stomping, and (8) a girl asking her father for ketchup. We included all descriptions of the targeted events from the eight videos, even when they did not contain syntactically identifiable quotation. Thus our corpus included canonical direct quotation constructions with quotatives and zero quotatives, and also *non-direct* reporting of the events. This latter category included indirect quotation (infrequent with non-speech) and descriptions, which could both be accompanied by bodily demonstration (see Examples 8–10; note that in example 8 the *like* is a discourse marker *like*, not a quotative *like*, Fox Tree, 2006).

(8) The kittens were like meowing to each other.
(9) They're talking about how they're going to make wine.
(10) It was just counting up from 1 to 10.

Thus the reporting or describing of any of the 8 events listed, with or without quotation devices, was included in the analysis.

Ratings. From the resulting corpus, for each clip we had about 35 direct quotations (using *be like*, *say*, and *go*, and zero quotatives) and 10 non-direct descriptions, depending on how many reporters included the critical event in their retellings. The quotations and descriptions were rated for the degree of vocal and bodily demonstration. The ratings ranged on a scale of 1 (not very demonstrative) to 5 (very demonstrative) and were made by research assistants who were naïve to the hypotheses of the study. To facilitate independent ratings across the two modalities, ratings for vocal demonstration were performed on the audio only, while bodily demonstration ratings were performed on the video without audio. Quotatives were excised in order to prevent bias.

As training, raters were first shown several example clips of participants' descriptions that illustrated a range of levels of vocal and bodily demonstration. Raters were instructed to first listen to or watch the full set of clips for a particular event in order to determine the range of demonstrations possible, and then to judge each retelling against the set. It was explained that vocal demonstrations included any way the speaker modulated his or her voice, deviating from the rest of the speech, in order to show something about the event. Bodily demonstration included any movements that could be thought of as using the body to show the event. Raters were instructed to take the whole body into consideration and to attend to both the size of movements and the number of movements produced. In order to not bias the raters, detailed instructions defining bodily demonstration were not provided, thus ratings could include postural movements, gaze movements, or facial expressions in addition to hand motions. Raters practiced until they reached a high level of inter-rater reliability with clips that were not used in the analyses. The actual clips were each rated by two raters (a total of six raters), with good inter-rater agreement for both vocal ($\boldsymbol{K}_w$ = .79, 95% CI, .75 to .83), $p < .001$) and bodily ($\boldsymbol{K}_w$ = .78, 95% CI, .74 to .82, $p < .001$) demonstration.

Because not all participants produced all five quotative devices, a mixed model linear regression was performed, with the device as a fixed factor, to determine the effect of the device type on the level of vocal and bodily demonstration.

2. Results

Table 1 gives examples of the vocal and bodily demonstrations produced by participants after viewing each of the original videos. For instance, when describing a video of a young girl asking her father for ketchup, participants often demonstrated her high-pitched voice to quote her request, ''Daddy pass the ketchup,'' often as they gazed upward as if toward the girl's imagined father. Or when describing a dog howling what sounds like ''I love you,'' participants howled the words, and like the dog, raised their head upwards in a howling motion.Table 2 shows the proportion of non-direct quotations found in the resulting corpus (from which the proportion of direct quotations can be derived by adding to 100%; for example, 19.2% of the Dog clip were indirect and 80.8% were direct). Table 2 also shows, within the direct quotations, the distribution of quotatives used (such that all direct quotations taken together add to 100%). Across all clips, the quotation device *be like* was used most frequently overall, 42.5%, followed by *say*, 22.6%, and *go*, 18.1%. Zero quotatives accounted for 6.2% of quotatives and *be just* for 3.9%. Other devices were infrequent, apart from some more specific reporting verbs associated with particular videos: *ask*, elicited by the ketchup video (e.g., ''the little girl asks, 'can I have some ketchup?'''') and *count*, elicited by the counting parrot (the parrots counts, ''one, two. . .'').

The distribution of quotatives differed from the distribution observed in a corpus of autobiographical narratives and conversations, where 92% of quotation devices were *like*, 7% were *said*, and 2% were *goes* (Fox Tree and Tomlinson, 2008; see also Blackwell and Fox Tree, 2012, who found a similar distribution with autobiographical narratives and narrative retellings). The clip-retelling technique increased the chances that a non-*like* quotative was used in direct quotations. Part of the reason for this could be because the clip-retelling technique encouraged quoting of a wider range of

Table 1
Examples of demonstrations produced in each of the clips.

Original video	Quoted speech/sound	Bodily demonstration
Talking dog	''I love you''	Raises head in a howling motion
Talking parrot	''I'm a double yellow-headed Amazon'' ''Grrr''	Bobs head
French accent	''Don't you know we French men are loyal?''	Clenches hand (like the speaker)
Grape stomping	''Ow, ow, ow''	Writhes in pain
Ketchup asking	''Daddy pass the ketchup''	Gazes upward as if to daddy
Babies babbling	''Da da da da''	Shakes arms (like the babies)
Counting parrot	''One, two, six''	Bobs head
Talking robot	''Let's go to the rocket ship''	Bends arms in a rigid position

Table 2
Percentages of quotative devices and non-direct quotation used for each stimulus video.

Clip	Type of sound	Direct quotations						Non-direct
		Be like	*Say*	*Go*	*Just*	Zero	Other	
Dog	Non-human speech	42.0	19.4	20.9	3.9	9.1	4.7	19.2
Talking parrot	Non-human speech	36.8	28.6	26.4	0	8.1	0	20.3
French	Human speech	40.3	32.8	15.4	2.6	6.3	2.6	42.2
Stomping	Human speech	68.1	10.6	11.7	4.3	3.2	2.1	24.6
Ketchup	Human speech	42.9	11.4	22.9	5.7	5.7	11.4	25.5
Babies	Human non-speech	34.1	10.7	28.6	9.7	9.7	7.1	44.4
Counting parrot	Non-human speech	11.1	55.6	11.1	0	0	22.2	34.3
Robot	Non-human speech	64	12	8	5.7	6.9	3.4	47.1
Overall		42.5	22.6	18.1	3.9	6.2	6.7	32.2

Table 3
Mean ratings (SD) of vocal and bodily demonstrations on a scale from 1 (not at all) to 5 (very).

	N	Vocal demonstration	Bodily demonstration
Zero	23	3.61 (.64)	3.94 (.83)
Be like	77	3.27 (1.10)	3.10 (1.13)
Go	68	3.31 (1.03)	3.07 (1.16)
Say	81	2.13 (1.25)	2.28 (1.15)
Non-direct	89	1.46 (.60)	1.69 (.78)

Table 4
Correlations between direct vocal and bodily demonstrations for each video.

Clip	Test statistics
Talking dog	$r(38) = .60$, $p < .001$
Talking parrot	$r(32) = .68$, $p < .001$
French accent	$r(34) = .73$, $p < .001$
Grape stomping	$r(35) = .80$, $p < .001$
Ketchup asking	$r(30) = .85$, $p < .001$
Babies babbling	$r(38) = .53$, $p = .001$
Counting parrot	$r(39) = .76$, $p < .001$
Talking robot	$r(43) = .55$, $p < .001$

behaviors than narratives and conversations. For example, if sounds are more likely to be introduced by *goes* (as suggested by Romaine and Lange, 1991), and sounds are more likely to be quoted in clip retellings (current study) than in autobiographical narratives, narrative retellings, and conversations (cf. Blackwell and Fox Tree, 2012; Fox Tree and Tomlinson, 2008), then there would be more quotations with *goes* in the current study.

Of the total 137 participants, 89 produced at least two different types of quotation devices and one non-direct quotation in the eight targeted videos. The mean ratings for vocal and bodily demonstration are given in Table 3.

A mixed model analysis, with device type (*be like*, *go*, *say*, zero, and non-direct) and demonstration type (vocal and bodily) as fixed factors, revealed a significant effect of device type on the level of demonstration, $F(4, 38.9) = 26.3$, $p < .001$. No difference was found between the demonstration types themselves, $F(1, 635.1) = 2.15$, $p = .07$, and no interaction was found between device type and demonstration type, $F(4, 635.1) = .52$, $p = .47$. For both vocal and bodily demonstration, the three highest ratings were for zero, *be like*, and *go*, then *say*, then non-direct. For vocal demonstrations, post hoc pairwise comparisons with Bonferroni adjusted levels of .005 (.05/10) showed similar ratings for zero, *be like* and *go* (no difference between zero and *be like*, $t(19) = .30$, $p = 77$, or between *be like* and *go* $t(56) = .85$, $p = .40$). There was, however, a reliable difference between *go* and *say*, $t(62) = 6.34$, $p < .001$, and between *say* and *non-direct*, $t(80) = 4.01$, $p < .001$. Similarly, levels of bodily demonstration were also higher for *be like*, *go*, and zero quotatives compared to *say* and non-direct quotations. Here, the level of demonstration was higher for *zero* than *be like*, $t(19) = 2.67$, $p = .02$, and higher than *go*, $t(19) = 2.83$, $p = .01$. There was no difference between *be like* and *go*, $t(56) = .16$, $p = .87$, but there was a difference between *be like* and *say*, $t(69) = .53$, $p = .001$, and between *say* and non-direct, $t(80) = 3.49$, $p = .001$.

Table 5
Correlations between vocal and bodily demonstrations for each quotative device and for non-direct quotation.

Type of quotation	Test statistics
Zero	$r(23) = .46$, $p = .003$
Be like	$r(77) = .54$, $p < .001$
Go	$r(68) = .55$, $p < .001$
Say	$r(81) = .44$, $p < .001$
Non-direct	$r(89) = .14$, $p = .29$

The relationship between the two types of demonstrations was then assessed for direct and non-direct quotation. In direct quotations, vocal and bodily demonstrations were strongly correlated, $r(249) .= .59$, $p .< .001$. Correlations for each video, with Bonferroni adjusted levels of .006 (.05/8), also revealed a positive association between these two types of demonstration for every clip (see Table 4).

For each device, with Bonferroni adjusted levels of .01 (.05/5), vocal and bodily demonstrations were also strongly correlated. They were not correlated for non-direct demonstrations (see Table 5).

3. Discussion

Quotation appears to be a multimodal construction. When speakers use direct quotations, they generally produce a high level of demonstration in both the vocal and bodily channels. Moreover, the level of demonstration in each channel is correlated. When speakers use more vocal demonstration, they also use more bodily demonstration when producing direct quotations. The highest amount of demonstration was found after *be like*, *go*, and zero quotatives. Of the direct quotation devices, *say* was associated with less vocal and bodily demonstration than the other three. Non-direct descriptions were the least demonstrative in the vocal channel, and, notably, were also the least demonstrative in the bodily channel.

The overall distribution of the quotatives obtained with the video-reporting task used here was different from the distribution found in narratives and conversations (cf. Blackwell and Fox Tree, 2012; Fox Tree and Tomlinson, 2008). The prevalence of *be like* was similar to the earlier observations. But at 18%, the use of *go* was considerably higher than the low percentage (2%) found earlier. One conclusion is that rather than falling out of use, as suggested by *like*'s becoming more popular and *goes*' becoming less popular over a twenty-year period (Fox Tree and Tomlinson, 2008), *go* may have become specialized for highly demonstrative quotations. Other evidence provided here suggests that *go* is especially useful for vocal demonstrations.

Across each of the four quotatives (*be like*, *go*, *say*, and zero quotatives), as well as across each of the videos, vocal and bodily demonstrations showed a strong positive correlation. While non-direct descriptions did not show this correlation, the overall level of demonstration was generally quite low for both channels. Previous studies have suggested a trade-off between gesture and speech under challenging conditions of speaking or gesturing, in which the speaker favors the easier mode of communication. In the present study, we considered the relationship of the vocal channel with gesture in the context of vocal demonstration, rather than speech. For this case, our findings support the Multimodal Hypothesis: a high correlation between activity in the vocal and bodily channels. When performing the demonstrative part of a quotation, the vocal and bodily channels are naturally integrated together into a unified, multimodal form. Indeed, previous work has shown that even in situations focused on sound, such as in the use of quotation in music instruction where physical motion has no literal meaning, gestures accompanying non-lexical vocalizations are still produced (Tolins, 2013). The lack of a correlation in non-direct speech could be a reflection of one manifestation of the Multimodal Hypothesis, with low vocal and low bodily demonstration.

We conclude with some broader consideration of the significance of quotation for investigating the relationship between language and gesture. A fundamental point of contention among cognitive scientists concerns whether language and gesture are complementary parts of a single system or separate systems of communication (e.g., McNeill, 1992; Levinson and Holler, 2014). Quotation offers a unique perspective into this relationship. In the formation of a quotation construction, the lexical selection of a quotative appears to be connected to the nature of the demonstration the speaker intends to perform. We found that some quotatives (e.g., *go* and *be like*) are associated with an increased overall level of demonstration in the following quoted content, whereas others (e.g., *say*) are associated with lower levels of demonstration. In this way, quotation bridges language and gesture, and connects the linguistic and imagistic processes that are involved in spoken communication. Moreover, the present study shows that quotation, particularly as it uses lexical cues like *be like*, *go*, and zero quotatives, is aptly considered multimodal, integrating the vocal and bodily channels into a single demonstration. Researchers have given considerable attention to how signed languages have the means to

describe an event through a grammatical integration of conventional morphemic forms and analogical demonstrations (e.g., classifier constructions; Emmorey, 2003; Liddell, 2003). Likewise, spoken languages may be seen to accomplish the integration of linguistic and analogical components through the use of the multimodal construction of quotation.

References

Andersen, Gisele, 1998. The pragmatic marker like from a relevance-theoretic perspective. In: Jucker, A.H., Ziv, Y. (Eds.), Discourse Markers: Descriptions and Theory. John Benjamins, Amsterdam, pp. 147–170.

Andersen, Gisele, 2000. The role of the pragmatic marker like in utterance interpretation. In: Andersen, G., Fretheim, T. (Eds.), Pragmatic Markers and Propositional Attitude. John Benjamins Publishing Company, Philadelphia, pp. 17–38.

Bangerter, Adrian, 2004. Using pointing and describing to achieve joint focus of attention in dialogue. Psychol. Sci. 15, 415–419.

Blackwell, Natalia, Fox Tree, Jean E., 2012. Social factors affect quotative choice. J. Pragmat. 44, 1150–1162.

Blyth, Carl, Recktenwald, Sigrid, Wang, Jenny, 1990. I'm like ''Say what?!'': a new quotative in American oral narrative. Am. Speech 65, 215–227.

Buchstaller, Isabelle, 2001. He Goes and I'm Like: The New Quotatives Re-visited Paper presented at New Wave 30, University of North Carolina.

Butters, Ronald, 1980. Narrative go ''say''. Am. Speech 55, 304–307.

Clark, Herbert H., Gerrig, Richard J., 1990. Quotations as demonstrations. Language 66, 764–805.

Coulmas, Florian, 1986. Reported speech: some general issues. In: Coulmas, F. (Ed.), Direct and Indirect Speech. Mouton de Gruyter, Berlin, pp. 1–433.

de Ruiter, Jan P., 2006. Can gesticulation help aphasic people speak, or rather, communicate? Adv. Speech Lang. Pathol. 8, 124–127.

de Ruiter, Jan P., Bangerter, Adrian, Dings, Paula, 2012. The interplay between gesture and speech in the production of referring expressions: investigating the tradeoff hypothesis. Top. Cognit. Sci. 4, 232–248.

Emmorey, Karen, 2003. Perspectives on Classifier Constructions. Lawrence Earlbaum and Associates, Mahwah, NJ.

Fox Tree, Jean E., 2006. Placing *like* in telling stories. Discourse Stud. 8 (6), 749–770.

Fox Tree, Jean E., Tomlinson Jr., John M., 2008. The rise of *like* in spontaneous quotations. Discourse Process. 45, 85–102.

Goodwin, Charles, 2007. Interactive footing. In: Holt, E., Clift, R. (Eds.), Reporting Talk. Cambridge University Press, Cambridge, pp. 16–46.

Hudson, Richard, 1985. The limits of subcategorization. Linguist. Anal. 15, 233–255.

Kendon, Adam, 2014. Semiotic diversity in utterance production and the concept of 'language'. Philos. Trans. R. Soc. Lond. B: Biol. Sci. 369, 20130293.

Levinson, Stephen C., Holler, Judith, 2014. The origin of human multi-modal communication. Philos. Trans. R. Soc. Lond. B: Biol. Sci. 369, 20130302.

Li, Charles, 1986. Direct speech and indirect speech: a functional study. In: Coulmas, F. (Ed.), Direct and Indirect Speech. Mouton de Gruyter, Berlin, pp. 29–45.

Liddell, Scott, 2003. Grammar, Gesture, and Meaning in American Sign Language. Cambridge University Press, Cambridge.

Mathis, Terri, Yule, George, 1994. Zero quotatives. Discourse Process. 18, 63–76.

McNeill, David, 1992. Hand and Mind. The Chicago University Press, Chicago.

Pascual, Esther, 2014. Fictive Interaction: The Conversation Frame In thought, Language, and discourse. John Benjamins Publishing Company, Amsterdam.

Romaine, Suzanne, Lange, Deborah, 1991. The use of *like* as a marker of reported speech and thought: a case of grammaticalization in progress. Am. Speech 66, 227–279.

Schiffrin, Deborah, 1981. Tense variation in narrative. Language 57, 45–62.

So, Wing C., Kita, Sotaro, Goldin-Meadow, Susan, 2009. Using the hands to identify who does what to whom: gesture and speech go hand-in-hand. Cognit. Sci. 33, 115–125.

Tannen, Deborah, 1989. Talking Voices: Repetition Dialogue and Imagery in Conversational Discourse. Cambridge University Press, Cambridge.

Tolins, Jackson, 2013. Non-lexical vocalizations in embodied enactments. Res. Lang. Soc. Interact. 46 (1).

van der Sluis, Ielka, Krahmer, Emiel, 2007. Generating multimodal references. Discourse Process. 44, 145–600.

Wade, Elizabeth, Clark, Herbert H., 1993. Reproduction and demonstration in quotations. J. Mem. Lang. 32, 805–819.

Waksler, Rachelle, 2001. A new all in conversation. Am. Speech 76, 128–138.

Yao, Bo, Scheepers, Christoph, 2011. Contextual modulation of reading rate for direct versus indirect speech quotations. Cognition 121, 447–453.

Yao, Bo, Belin, Pascal, Scheepers, Christoph, 2011. Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. J. Cognit. Neurosci. 23 (10), 3146–3152.