Research Article   **Open Access**

Kashmiri Stec*, Mike Huiskes, Gisela Redeker

Multimodal analysis of quotation in oral narratives

Abstract: We investigate direct speech quotation in informal oral narratives by analyzing the contribution of bodily articulators (character viewpoint gestures, character facial expression, character intonation, and the meaningful use of gaze) in three quote environments, or quote sequences – single quotes, quoted monologues and quoted dialogues – and in initial vs. non-initial position within those sequences. Our analysis draws on findings from the linguistic and multimodal realization of quotation, where multiple articulators are often observed to be co-produced with single direct speech quotes (e.g. Thompson & Suzuki 2014), especially on the so-called left boundary of the quote (Sidnell 2006). We use logistic regression to model multimodal quote production across and within quote sequences, and find unique sets of multimodal articulators accompanying each quote sequence type. We do not, however, find unique sets of multimodal articulators which distinguish initial from non-initial utterances; utterance position is instead predicted by type of quote and presence of a quoting predicate. Our findings add to the growing body of research on multimodal quotation, and suggest that the multimodal production of quotation is more sensitive to the number of characters and utterances which are quoted than to the difference between introducing and maintaining a quoted characters' perspective.

Keywords: quotation, multimodality, viewpoint

1 Introduction

Narratives – personal and otherwise – are pervasive in human interaction (Turner 1998; Zunshine 2006). They engage our imagination by presenting real or fictional events outside the here-and-now of the story telling, often by enacting (or, in Clark & Gerrig's 1990 terms, "demonstrating") the characters' speech or thought in direct quotation (Clark & Van Der Wege 2001). Direct quotation shifts the deictic center (time, space, personal pronouns) to the original utterance or thought, thus assuming the quoted character's perspective and shifting viewpoint to them (Dancygier & Sweetser 2012). Consider the following extract from our collection of informal oral narratives. The narrator, Black, describes a last-minute costume change she requested as a child on Halloween.

(1) Black: or maybe I didn't want to be the clown I don't know
        [past.self] *mom I don't want to be a clown*
        [mother]   *well um it's kind of last minute*
                    all the Halloween stores are closed

In this example, Black quotes a dialogue between her past self and her mother. The tense shifts from past to present, but there are no other linguistic cues such as *say* or *think* which might prepare the listener

*Corresponding author: Kashmiri Stec:** Center for Language & Cognition, University of Groningen, E-mail: k.k.m.stec@rug.nl
Mike Huiskes, Gisela Redeker: Center for Language & Cognition, University of Groningen

for the viewpoint shifts which occur as the quoted dialogue unfolds. As we shall see in a more detailed discussion of this example (in §3.3), Black uses special prosody for each quoted character, looks away from her addressee in slightly different directions (left for her past self; left and down for her mother) and produces character-specific facial displays as well – a disappointed look when quoting her past self, and a frustrated look when quoting her mother. In this study, we investigate how such multimodal cues are used to signal the perspective shifts involved in quoting speech and thought in informal oral narration.

1.1 Linguistic realization of quotation

The structural and interactional functions of quotation have attracted considerable attention (e.g. Banfield 1982; Li 1986; Tannen 1989; Redeker 1991; Dancygier & Sweetser 2012; Yao et al. 2012; Eerland et al. 2013; Groenewold et al. 2014; Guerini 2014; Pascual 2014, among many others). This interest has stemmed from two observations (1) quotation is ubiquitous in discourse, whether in healthy speaking (Banfield 1982; Tannen 1989; Pascual 2014), healthy signing (Engberg-Pedersen 1993; Metzger 1995) or brain-damaged speaking populations (Berko-Gleason et al. 1980; Hengst et al. 2005; Groenewold et al. 2014), and (2) despite appearances, quotation does not commit the speaker to the veracity of either the form or the content of the quoted utterance (Bakhtin 1981; Clark & Gerrig 1990; Redeker 1991; Sams 2010). Instead, the interactional function – i.e. what purpose the speaker is trying to achieve with the quoted utterance – takes precedence.

Understanding quoted utterances relies on our ability to fluidly adopt, represent and understand multiple perspectives in ordinary discourse. This is in part because quotation makes the story lively (Groenewold et al. 2014) or vivid (Li 1986; Mayes 1990) by creating involvement (Tannen 1989), dramatizing interaction (Labov 1972, Redeker 1991) and stimulating our imagination (Clark & Van Der Wege 2001). At the same time, quoting speakers are often more committed to preserving the intended meaning of a quoted utterance than its form (Lehrer 1989; Redeker 1991; Eerland et al. 2013). They may also use quotation without any 'reporting' function at all, e.g. when voicing non-human entities (the dog was howling "*feed me!*"), inanimate objects (Firefox was like "*Nope, you can't see that webpage*"), or thoughts and attitudes (he was all "*Boring!*"). In these situations, speakers quote utterances which were never actually uttered – whether in form or content. Such uses have been variously termed *constructed dialogues* (Tannen 1989), *dramatizing quotes* (Redeker 1991), *enactments* (Streeck & Knapp 1992), *hypothetical active voicing* (Simmons & LeCouteur 2011) or *fictive interaction* (Pascual 2014).

Because speakers use direct quotation outside of reporting contexts and for different interactional purposes, researchers have argued that it would be more appropriate to say that speakers *demonstrate* (Clark & Gerrig 1990) or *reenact* (Sidnell 2006) rather than describe or reproduce previous actions and events, and that they do so "by creating the illusion of the listener also being an eye-witness" (Sakita 2002: 189). Direct speech constructions are particularly suited to this, as they help animate and enact characters in a narrative (Goodwin 2007) and serves higher-order discourse functions, such as acting as an objective evidential outside of narrative contexts (Couper-Kuhlen 2007).

1.2 Multimodal realization of quotation

In addition to linguistic indicators of quotation, which are produced in the acoustic modality, narrators may use co-produced paralinguistic features or multimodal bodily actions. Paralinguistic features which are important for indicating a shift to a quoted perspective are changes in intonation or prosody (Couper-Kuhlen 1999; Couper-Kuhlen 2007) or rate of speech (Yao et al. 2012). However, the visual modality can also contribute in meaningful ways. Most work on visible bodily actions accompanying speech focuses on contributions made by the hands in different communicative situations, such as problem solving (e.g. Holler & Wilkin 2009), telling narratives (e.g. Parrill 2012), or face-to-face interaction (e.g. Kendon 2004). Within the field of gesture studies, perspective is typically studied via manual gesture production, such as when a speaker produces a *character viewpoint gesture* (e.g. pumping both arms to demonstrate "running") rather than an *observer viewpoint gesture* (e.g. moving an index finger in front of the speaker's body to demonstrate "running") (McNeill 1992).

However, as even early gesture work points out, the body is composed of multiple articulators, and people are adept at using those articulators in meaningful ways. Thus, as with sign languages (e.g., Cormier et al. 2012), manual co-speech gestures may be co-produced with other articulators in the visual modality, such as facial expression (Chovil 1991), changes in gaze (Sidnell 2006; Park 2009; Thompson & Suzuki 2014), or the use of gestural space (Özyürek 2002; Perniss & Özyürek 2015). In narrative contexts, Earis & Cormier (2013) find that multimodal character representations in English often include prosodic elements, character facial displays and the use of co-speech gestures (though not necessarily character viewpoint gestures) to "enrich the narrative discourse" (p.339). Stec (2012: p.351-353) provides a summary of the articulatory means by which viewpoint shifts may be indicated by the gesturing body; Cormier et al. (2013) present a similar analysis for sign. In brief: People will create and construe meaningful differences using whatever means are available.

Considering in particular the visible bodily actions accompanying quotation, a number of qualitative studies have noted that multiple multimodal articulators are used in coordination with speech (e.g. Clark & Gerrig 1990; Streeck & Knapp 1992; Sidnell 2006; Buchstaller & D'Arcy 2009; Park 2009; Fox & Robles 2010), with certain contexts of use garnering specific multimodal production strategies. One such example comes from Park (2009), who suggests that Korean speakers systematically use different multimodal behaviors depending on whether they use a direct speech construction to quote past-self (present), past-addressee (present) or a past third person character (absent), even though the Korean quoting particle already makes this distinction. Gaze is an important articulator for managing interaction in the here-and-now vs. representing other interactions (Sweetser & Stec in press), and can be used to indicate whether an interlocutor is being quoted (Park 2009) or is 'standing in' for the addressee of a prior exchange (Thompson & Suzuki 2014). Facial expressions may also be used when describing characters in a narrative (McNeill 1992; Sidnell 2006), or to differentiate between a described scene and the present moment (Chovil 1991). Intonation and other prosodic or phonetic changes may be used not only to demonstrate what someone else said (Clark & Gerrig 1990; Fox & Robles 2009), but also to differentiate between the speaker and quoted characters (Couper-Kuhlen 1999; Couper-Kuhlen 2007) or even to make meta-comments on the quoted utterance (Günthner 1999).

1.3 Research questions

Previous research has demonstrated that there are different uses of quotatives in discourse, and different multimodal realizations accompanying them. These discussions typically focus on single quoted utterances, such as *The President said "We're going to Mars!"* However, as example (1) above shows, there are also different quotation environments, or *quote sequences*. Our notion of quote sequence is intended to capture the fact that while single quoted utterances are common (Buchstaller 2013), speakers can quote in dialogic or monologic sequences in addition to the single quotes which are often studied. In other words, there are several possibilities for the production of quotations: (1) there is only one quoted utterance, immediately after which the narrative or interaction with interlocutor(s) resumes (single quotes, or *Quote Islands*), (2) there are multiple quoted utterances, all from the same quoted speaker (*Quoted Monologues*), or (3) there are multiple quoted utterances, from multiple quoted speakers (*Quoted Dialogues*).

This distinction is important for two reasons. First, the number of quoted speakers could affect the multimodal strategies used. For example, when adopting character perspective in sign languages, there is evidence that the strategies used to enact two or more characters with contrastive roles (Padden 1986) differ from the strategies used to enact characters elsewhere (Janzen 2012). As speakers and signers share certain capacities for iconic representation (Perniss et al. 2015; Vigliocco et al. 2014), it might be the case that speakers use different multimodal articulators when a different number of characters are involved. Second, the number of quoted utterances could affect the multimodal strategies used. As several studies have demonstrated, not only do referential introduction and maintenance contexts garner different multimodal behaviors in both speakers (Levy & McNeill 1992; Perniss & Özyürek 2015) and signers (Cormier et al. 2012; Perniss & Özyürek 2015), speakers appear to preferentially mark the introduction of quoted utterances with multimodal articulation (Sidnell 2006). It might therefore be the case that the initial utterance of a multi-

utterance quotation sequence is produced with different multimodal articulators or actions compared to continuing utterances, where quoted interaction is maintained.

In this paper we report a corpus-analytic investigation of the role of multimodal indicators of perspective change and character embodiment in the production of direct quotation in oral narratives. We focus on the use of speech- or thought-report predicates (*he says*, *I'm like*, *I thought*, etc.), type of quotation (direct speech vs. fictive interaction) and a set of bodily articulators which may signal a shift to character viewpoint through the use of intonation, facial expression, gaze, or manual gestures which are used to demonstrate certain aspects of the quoted character. We expect short, single-utterance quotations (Quote Islands) to be treated differently from quoted sequences. Quoted Dialogues involve more than one character's perspective and are thus expected to attract more multimodal indications of perspective shift than quoted sequences of utterances from a single character (Quoted Monologues). Finally, we expect initial utterances to be marked with more multimodal indicators of perspective shift than non-initial utterances.

The remainder of this paper is structured as follows: In Section 2 we describe our dataset and the means by which it was annotated. In Section 3 we present a qualitative description of multimodal quotation in our dataset, focusing on QIs (§3.1), QMs (§3.2) and QDs (§3.3). In Section 4 we present a quantitative analysis of multimodal quotation, focusing first on the frequencies of multimodal production in quotation sequences (§4.1) and then on modeling quotation sequences with linguistic and multimodal predictors (§4.2). Finally, we discuss our results in Section 5, and situate them within a broader discussion of multimodal quotation and multimodal character viewpoint.

2 Method

We conducted an exploratory analysis on a corpus of semi-spontaneous speech which was collected and annotated by the first author. We give a brief overview of the collection and annotation procedure below; more detail can be found in Stec et al. (under review). A paper package containing relevant methods documents (codebook, data, R code, etc.) is available for download at the Mind Research Repository.

2.1 Data

We collected semi-spontaneous narratives from pairs of native speakers of American English. Participants volunteered their time, and completed a 2-step consent procedure in which they first consented to participate in data collection and later granted specific use of the materials which had been recorded (e.g. publication of the stills in Section 3). All pairs of speakers knew each other, and were asked to tell each other personal narratives they would be comfortable having on film. Some dyads improvised, others used an optional topic sheet which had been prepared by the first author to guide their interaction. Our corpus consists of 26 speakers (19 female) and 85 narratives which range in length from 0:30 to 15:51 minutes, with an average length of about 5 minutes per narrative. There are 704 quoted utterances in our corpus.

2.2 Annotation

Our analyses are based on annotations made by the first author on the entire corpus. To obtain interobserver validity of annotations, we used a consensus procedure whereby annotations made by the first author on a subset of the corpus (10%) were compared against annotations made by four independent coders in successive rounds of annotation. These comparisons were facilitated by lengthy discussions of the annotation choices made until the underlying sources of disagreement were identified and resolved. We did not measure agreement by Kappa, as such measures can mask the underlying source of disagreement (Stelma & Cameron 2007) or erroneously indicate agreement (e.g. Uebersax 1987; Gnisci et al. 2013). More information about our consensus procedure is available in Stec et al. (under review).

The relevant portion of our final annotation scheme is presented in Table 1. This scheme was implemented in ELAN, video annotation software developed by the Max Planck Institute for Psycholinguistics – see

Wittenburg et al. (2006) and http://tla.mpi.nl/tools/tla-tools/elan/ for more information. Our annotation scheme includes variables (tiers) for linguistic features pertaining to quotatives and for multimodal articulators which have been identified as relevant for the multimodal expression of viewpoint in studies investigating speaking populations (viz. Sidnell 2006; Park 2009; Earis & Cormier 2013 – see Stec 2012 for a review).

Table 1: ELAN tiers and controlled vocabularies used (adapted from Table 3 in Stec et al. under review).

Category	Tier	Controlled Vocabulary
Linguistic Information	Transcript	*Text*
	Quote type	Direct speech (the utterance voices a character)
		Fictive interaction (the utterance represents a character's thoughts, voices an entity which cannot speak, or refers to a future, pretend or counterfactual scenario, cf. Pascual 2014)
		Unclear
	Quote sequence type	Quoted dialogue (two or more characters are quoted over successive utterances)
		Quoted monologue (one character is quoted over several utterances)
		Quote island (one quoted utterance)
		Other
	Quote position	Initial (first quoted utterance in a multi-utterance sequence or quoted utterance that is a Quote Island)
		Continuing (non-initial quoted utterance)
Multimodal Articulators	Character Intonation	Present (speaker's voice altered to demonstrate the quoted character)
		Absent (speaker's voice unchanged)
		Unclear
	Hands	Character viewpoint gesture (speaker's hands demonstrate a manual action performed by another entity)
		Other gesture (including beats, iconic gestures which are not character viewpoint, deictic gestures, emblems, etc.)
		No gesture
	Character Facial Expression	Present (speaker's facial expression changes to demonstrate the quoted character)
		Absent (speaker's facial expression is unchanged)
		Unclear
	Gaze	Maintained with addressee (speaker's gaze is directed to addressee throughout the quote)
		Away from addressee (speaker's gaze is not directed to the addressee throughout the quote)
		Late change (speaker's gaze moves away from the addressee after the quote started)
		Quick shift (speaker's gaze jumps around throughout the quote)
		Unclear
Notes	Notes	*Notes*

2.2.1 Linguistic features

We first noted whether an utterance was a quote, and whether that quote was an instance of fictive interaction (Pascual 2014). Quotes were identified by the presence of a quoting verb (*be like, say, think*, etc.) and, in the case of bare quotes, by the use of direct speech (such as *I* instead of *he*, a switch to present tense or other deictics, and so forth – see Buchstaller 2013 for more information about identifying direct speech in discourse). We also noted if the utterance was spoken with special intonation -- that is, whether it differed in any way from the speaker's normal voice (henceforth, *character intonation*). This might be when a narrator whispers or shouts to demonstrate how a character actually spoke, or any change in pitch or loudness compared to the speaker's normal voice. What was important to us was that there was an observable difference, not what the particular phonetic realizations of that difference might be.

2.2.2 Quote sequences

To annotate the quote environment of each quoted utterance, we first noted whether it was an instance of a particular quote sequence: a Quote Island (QI), a Quoted Monologue (QM), or a Quoted Dialogue (QD). Nine utterances (1.3% of the dataset) were coded as 'Other' and were excluded from this study.[1] Our dataset thus consists of 695 utterances. An example of each quote sequence is given below; further discussion of these examples appears in Section 3. Throughout our paper, examples are formatted as follows, Speaker_Name: [quoted.speaker] *quote*.

   QIs are single quoted utterances by a single speaker.

(2) Quote Island
   Grey: [Picasso] *darn kids*

QMs are multiple quoted utterances in a row by the same speaker, with only quoting predicates appearing between the quoted utterances (e.g. *and she said* or bare quotes) .

(3) Quoted Monologue
   Pink: [past.self] and I'm like *oh this must be a special one really*
                  just for my headset
      [past.self] and I'm like *oh it's not working*

QDs are multiple quoted utterances in a row by different quoted speakers, with only quoting predicates appearing between the quoted utterances. In our dataset, QDs only occur between two quoted characters, e.g in example (4) the speaker's past self and her mother.

(4) Quoted Dialogue
   Black: or maybe I didn't want to be the clown I don't know
         [past.self] *mom I don't want to be a clown*
         [mother] *well um it's kind of last minute*
                  all the Halloween stores are closed

As these examples demonstrate, quote sequences differ with respect to the number of quoted speakers (1 or 2) and the number of quoted utterances (QIs – 1, QMs – 2 or more, QDs – 2 or more).

   In addition to quote sequence, we coded all utterances as being initial (i) or continuing (c) in a given quote sequence. All QIs were coded as initial. If a quote was identified as a QM, we annotated the first quote in the sequence as initial, and all other quotes in the sequence as continuing. If a quote was identified as a QD, we annotated the first quote as "initial A" and the first quote by the other speaker as "initial B"; all other quotes in the sequence were annotated as continuing. An example of a coded quote sequence is given in (5):

(5) Pink: [past.self] and I'm like *oh this must be a special one really*
                  just for my headset   [QMi]
         [past.self]and I'm like *oh it's not working*   [QM2/c]

In this example, the first utterance in the sequence (*oh this must be a special one really just for my headset*) is coded as QMi. This code identifies the utterance as both being part of a Quoted Monologue (QM) and as the initial utterance in the QM sequence (i). The second utterance (*oh it's not working*) is coded as QM2/c. This code identifies the utterance as being the second (2) or non-initial/continuing (c) utterance in a QM

---

1 'Other' quotes came in four types: (i) echos, the addressee repeats the quote the speaker just produced (3 quotes); (ii) completions, the addressee completes the quote the speaker starts (1 quote); (iii) generics, of the form *he said that and she said that* (2 quotes); and (iv) interruptions, which occur in one narrative when a quoted sequence is interrupted by the addressee (3 quotes).

sequence. All initial utterances were the first utterances in their sequence; the number of continuing utterances ranged from 2 (48 QM sequences and 38 QD sequences) to 19 (1 QD sequence).

2.2.3 Multimodal articulators

Although most co-speech gesture research focuses on the expressive capacities of the hands (e.g. McNeill 1992 and Kendon 2004, among many others), multiple articulators contribute to multimodal utterances. Several qualitative studies have emphasized this point (e.g. Sidnell 2006; Thompson & Suzuki 2014), and quantitative work is starting to do so (e.g. Earis & Cormier 2013). Stec (2012: p. 351-353) provides a summary of the use of different multimodal articulators in conveying multimodal viewpoint in co-speech and co-sign gesture. The use of gaze, facial expression, manual gestures and the use of gesture space play important roles in the expression of *conceptual viewpoint*, i.e. the real or imagined physical location of a conceptualizer. The notion of conceptual viewpoint can be used to capture differences between quoted characters, such as when a speaker reproduces a past interaction and demonstrates their interlocutors' actions and their own past actions (example 4 above), or the difference between observer or character perspective when describing an event (Parrill 2012). Knowing which articulators can be used to indicate conceptual viewpoint shift (such as when adopting character perspective in narrative contexts, cf. Earis & Cormier 2013) can help to identify the articulators which are specifically used to demonstrate character viewpoint during quotes.

We annotated each of the features identified by Stec (2012), focusing on articulators which "actively" contribute to the multimodal expression of viewpoint. As an example of what we mean by "active" articulator use, consider a character viewpoint gesture which is produced with a display of affect on the face versus one which is not, such as a speaker pumping their arms to show a character running with or without a co-timed display of fear on the speaker's face. Character viewpoint gestures use the speaker's body to show how a character acted. According to McNeill (1992), character viewpoint would be mapped onto the speaker's entire body regardless of whether or not a facial display of character affect is present. However, in the first instance (gesture with a fearful expression), we would say that both the hands and the face are "active" as both contribute to the speaker's representation of the character while in the second instance, only the hands are "active" and therefore only the hands contribute to the speaker's representation of the character. We are only interested in "active" uses of multimodal articulators.

We annotated three nonverbal articulators for their "activation": gaze, facial displays, and manual gesture. For gaze, we identified four kinds of behaviors: the speaker's gaze is maintained with the addressee, moves away from the addressee at the start of the quote, moves away a few words into the quote (late change), or is undirected and jumps all over the place during the quote (quick shift). For the purpose of modeling (§4.2), we identified the latter three behaviors (looking away, late change and quick shift) as being indicative of viewpoint shift and treated them as a group.[2] Henceforth, we refer to them jointly as "the meaningful use of gaze". We noted whether the speaker's face displayed the affect of the quoted character, e.g. with facial expressions attributable to the character being quoted. For manual gesture, we made a three-way distinction between character viewpoint gesture, other gesture, and no gesture. This simplified scheme reflects the fact that our notion of "activity" is specifically linked to character viewpoint, and the fact that speakers have been shown to use similar behaviors as signers when adopting character perspective, such as the combined use of manual character viewpoint gestures with character facial displays (Liddell & Metzger 1998; Earis & Cormier 2013) or the systematic use of gaze to differentiate viewpoint (Sweetser & Stec, in press).

---

2 As one reviewer pointed out, speaker gaze is complicated (see, e.g., Thompson & Suzuki 2014 or Sweetser & Stec in press) and our variable simplifies it. Our decision to count "not maintaining gaze with addressee" as "meaningful use of gaze" was based on Sidnell (2006), who found that speakers often look away from interlocutors on the left boundary of quotes. This is similar to what some signers have been reported to do during constructed action or role shift sequences (e.g. Metzger 1995). However, as the reviewer rightfully pointed out, a speaker who consistently gazes away from their addressee during the narrative but who consistently gazes towards their addressee when quoting is "meaningfully" using gaze, but this is not captured by our coding scheme.

3 Multimodal quote sequences

3.1 Quote Islands

When Quote Islands (QI) occur, they are single quoted utterances which are preceded and followed by other linguistic actions – asides to the interlocutor, other aspects of the narrative, etc. They voice only one character. There are 298 QI sequences (and thus, 298 QI utterances) in our dataset.

An example is given in Transcript 1 and Figure 1. Each line of the transcript corresponds to an image with that number in the figure; i.e. line 1 of the transcript was produced with the actions shown in Figure 1, image 1. In this example, White, the narrator of this story, appears on the left of the frame. Her addressee Grey appears on the right. In this narrative, White describes a visit to a museum as a small child. Prior to the museum visit, White had been impressed by Picasso's work, but after seeing the exhibit, she felt unimpressed and ends the narrative in line 1 by saying that Picasso is probably *rolling in his grave* while producing an observer-viewpoint gesture which shows three repetitive rolling actions. This is shown in Figure 1, image 1. Her addressee, Grey, mimics her observer-viewpoint gesture but with slower, larger movements which include his upper torso, head and visible beats each time he completes a circle during the pause on line 2 of the transcript (Figure 1, image 2). He uses character intonation while uttering the QI in line 3 and scrunches up his face, as if in a scowl. This utterance is an instance of fictive interaction since Grey is voicing something which Picasso, long dead, couldn't have said in response to this particular situation. While producing line 3, Grey looks away from his addressee and moves his head sideways (Figure 1, image 3). Grey returns to maintaining White's gaze in line 4 and his body comes to rest, signaling to White that she can resume her narrative.

Figure 1: Stills from *Mirrors*.

Transcript 1: *Mirrors* (White's narrative)

1 White: he's probably like rolling in his grave
2   (0.5s)
3 Grey: [Picasso] *darn kids*
4   or I don't know something um French

Multiple multimodal articulators contribute to this Quote Island: Grey uses character intonation, a character facial expression (the scowl), a change in gaze and head movements. [3]

3.2 Quoted Monologues

In Quoted Monologues (QM), one perspective is quoted multiple times in a row, with no intervening linguistic actions apart from quoting predicates (e.g. *she said* or bare quotes) or extended quoting predicates, such as *and then she was all*. Often a character is only quoted twice, but in some cases, three or more successive quotes occur. There are 59 QM sequences in our dataset, with a total of 138 QM utterances. Most QM sequences are comprised of two utterances (48 sequences), with a range of up to seven utterances (one sequence). An example of a QM is given in Transcript 2 and Figure 2, which is taken from Pink's narrative about a visit to a museum in Korea.
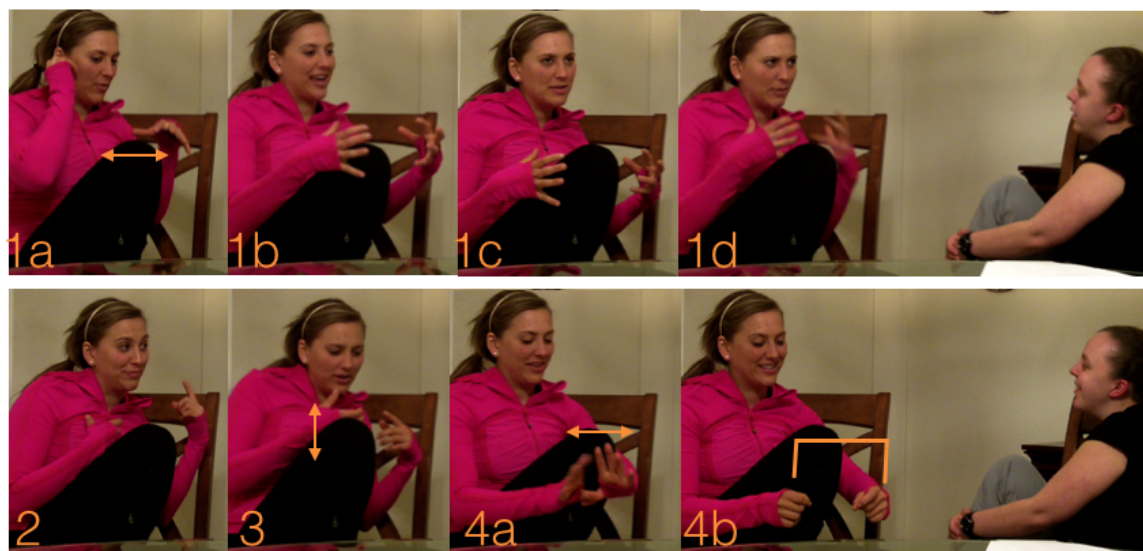

Figure 2: Stills from *Korea*.

Transcript 2: *Korea* (Pink's narrative)

```
            (1a)            (1b)        (1c)              (1d)
1 Pink: it doesn't have any words it just shows like headset
2       [past.self] and I'm like oh this must be a special one really
                    just for my headset
3       [past.self] and I'm like oh it's not working
                    (4a)                          (4b)
4       meanwhile next to the headset next to the like image of the headset
```

Pink appears on the left of the images, and the QM sequence is italicized on lines 2 and 3 of the transcript. When this example begins, Pink's body is oriented towards her narrative space, which is situated between herself and her addressee. She uses her hands to shape the museum exhibit in front of her (line 1) using observer-viewpoint gestures, as well as her failed attempts to get her own headset, with a spoken English guide to the museum's exhibits, to interact with it (lines 2 and 3). Figure 2, images 1a-d show the gestures

---

3 Although not considered here, head movements are also a rich source for the management of discourse, and transitioning between different characters within narratives – see McClave (2000) for further details on the multifunctionality of head movements.

accompanying line 1 of the transcript: Pink first uses a sweeping gesture (image 1a) and then an observer-viewpoint gesture (images b-d) to depict the museum exhibit. In line 2, the first QM utterance, her left hand points to her head (Figure 2, image 2). She uses special character intonation, looks away from her addressee at the start of the quote, and displays surprise on her face. While uttering line 3, the second QM utterance, Pink's right hand, now holding the imagined headset (a character-viewpoint gesture), moves up and down depicting her attempts to get it to work, while her left hand returns to rest (Figure 3, image 3). Like the previous QM utterance, she uses special character intonation, looks away from her addressee, and displays surprise on her face. The narrative continues in line 4 as Pink starts to trace a piece of the exhibit (Figure 2, images 4a-b). Throughout this sequence, her head alternates from looking to her gestures and her addressee, her gestures iconically depict features of the described scene, and her face expresses the surprise and confusion she felt at failing to get her own headset to interact with the exhibit. Here, too, we see multiple multimodal articulators working together in the production of the quotes.

3.3 Quoted Dialogues

In Quoted Dialogues (QD), the narrator quotes multiple characters in immediate succession, often reporting a dialogue which actually took place. There are no intervening linguistic actions between QD utterances apart from quoting predicates (e.g. *she said* or bare quotes) or extended quoting predicates, such as *and then she was all*. Most QDs are introduced without a quoting predicate. There are 76 QD sequences (213 QD utterances) in our dataset. There are often two utterances (38 QD sequences) or three utterances (24 QD sequences) per QD, but one narrative consists almost entirely of two QD sequences (21 and 16 quoted utterances; see Stec et al. under review for more about this narrative).[4]

An example of a QD is given in Transcript 3 and Figure 3, which is taken from a narrative about Black's failed attempts to find a suitable Halloween costume as a small child. Black is on the right of the frame and her addressee, Pink, is on the left. The QD is italicized on lines 2 and 3 of the transcript: in line 2, Black voices her past self, and in line 3, she voices her mother's response.

Figure 3: Stills from *Halloween*.

---

4 QD sequences in our dataset follow two patterns: quoted speakers alternating at every utterance (an A B… pattern) or not (an A A B…). Only four QD sequences follow the second pattern, and are excluded from the analyses presented in Tables 7, 11 and 12 in §4.2.

Transcript 3: *Halloween* (Black's narrative)

1 Black: or maybe I didn't want to be the clown I don't know

2   [past.self] *mom I don't want to be a clown*

                    (3a)   (3b)

3   [mother] *well um it's kind of last minute all the Halloween stores*

                are closed

4 Pink: be the clown and *eat your tea*

5 Black: *eat your tea*

     (6a)   (6b)

6   so I didn't want to be the clown

When the example begins on line 1, Black is looking at her addressee, Pink (Figure 3, image 1). Black then orients towards her narrative space, which is situated slightly to the left of the space shared with her addressee. While voicing her past self in line 2, she looks to her left while using special intonation for herself as a child (Figure 3, image 2). Simultaneously, she also produces a facial expression which evokes disappointment. When Black voices her mother in line 3, she looks down and to the left, and uses a character-viewpoint gesture to express her mother's frustration (Figure 3, images 3a-b). She also uses special frustrated character intonation for her mother, and produces a character facial display which evokes that frustration. When Pink interjects with a fictive quote in line 4, Black shifts her head and gaze towards her interlocutor and looks down while simultaneously echoing the second half of Pink's interjection, *eat your tea* (Figure 3, image 4-5)[5]. The near simultaneity is indicated in the transcript by the * on either side of this phrase. Black then resumes her narrative in line 6 (Figure 3, images 6a-b).

For this Quoted Dialogue, we see different multimodal resources being used for each quoted character: Black's past self is represented with fewer active articulators – character intonation and facial expression, and a change of gaze – while Black's mother is represented with more active articulators – character intonation and facial expression, character viewpoint gesture, and a change of gaze. Although this is a difference of only one articulator (manual gesture), the multimodal activity accompanying Black's past self is minimal in comparison, since it both uses less of her gesture space and the active articulators are less distinct.

In this section, we have seen that in each of the three quoted contexts (QI, QM, QD) multiple multimodal articulators accompany quoted utterances. These articulators may be, e.g., the use of character intonation or facial expression, the meaningful use of gaze, and the use of gesture or body movement in addition to speech. Our findings are consistent with previous work which found extensive use of character intonation, facial expression and gaze when quoting or enacting characters (e.g. Sidnell 2006; Earis & Cormier 2013; Thompson & Suzuki 2014) but limited use of manual character viewpoint gestures (Earis & Cormier 2013). To this body of work, we add our observations about multimodal QD sequences, where, like Padden (1986) observed for sign language, contrastive behaviors may be used to distinguish multiple characters.

4 Results

Our analysis comprises two parts: In Section 4.1 we present the frequency of multimodal articulator use in each quote sequence, and in Section 4.2 we attempt to model quote sequences on the basis of multimodal articulation (§4.2.1) or utterance location within a sequence (§4.2.2). If speakers manage character perspectives via multimodal production, then we expect to find unique sets of multimodal predictors for single-character quote sequences (QIs and QMs) on the one hand and multiple-character quote sequences

---

5 This phrase (*eat your tea*) was adopted by Pink and Black early in their recording session as a humorous way of expressing alignment. No character in any of their narratives actually utters this phrase. Rather, it is used fictively, in the sense of Pascual (2014), to demonstrate that the listener is actively following the speaker's story. In this case, Pink uses the fictive utterance to demonstrate what Black's mother would not have done – i.e., insist that young Black wear an unwanted Halloween costume.

(QDs) on the other. We might also expect to find a difference between single characters which are quoted once (QIs) or over several utterances (QMs). If, however, the articulation is sensitive to changes in viewpoint from the narrator to a quoted character, then we expect to see a reduction in the number of multimodal predictors co-occurring with continuing compared to initial utterances across quote sequences (QI + QMi + QDi vs. QMc + QDc) and within quote sequences (QMi vs. QMc and QDi vs. QDc).

4.1 Frequency of multimodal articulator use in quote sequences

We begin with an overview of the multimodal features accompanying quote sequences in our dataset. Because the distinction between initial and non-initial utterances is important for the regression results presented in Section 4.2, we make a distinction between QIs, initial and continuing QMs and initial and continuing QDs.

In Table 2, we see the distribution of quoted and fictive speech in the three types of quote sequences. Overall, there are twice as many direct speech utterances (66.5%) as fictive interaction utterances (33.5%) in the corpus. Most utterances are QIs (49.2%), and there are almost twice as many QDs (37.2%) as QMs (19.9%). Each quote sequence differs with respect to the type of quoted speech that occurs: while QIs consist of somewhat more direct speech utterances (25.8%) than fictive interaction utterances (17.1%), QDs overwhelmingly consist of direct speech utterances (31.4%) with very few fictive interaction utterances (5.9%) and QMs consist of almost equal amounts of direct speech utterances (9.4%) and fictive interaction utterances (10.5%). In both QMs and QDs, there are more fictive interaction utterances in non-initial than initial position within the quote sequence: 6.2% vs. 4.3% for QMs and 4.5% vs. 1.4% for QDs.

Table 2: Quoted and fictive speech in the three types of quote sequences

Quote Sequences	Quoted Speech		Fictive Speech		Total	
	N	%	N	%	N	%
Quote Island	179	25.8	119	17.1	298	42.9
Quoted Monologue initial	29	4.2	30	4.3	59	8.5
Quoted Monologue continuing	36	5.2	43	6.2	79	11.4
Quoted Dialogue initial	66	9.5	10	1.4	76	10.9
Quoted Dialogue continuing	152	21.9	31	4.5	183	26.3
Total	462	66.5	233	33.5	695	100

The distribution of quoting predicates is given in Table 3. There are nearly twice as many uses of quoting predicates with verbs like *say* or *think* (69.8%) as there are uses of bare quotes (30.2%). QIs are generally introduced by a quoting predicate (79.5%). QM initials are more often introduced by a quoting predicate (78%), while QM continuing utterances are more often introduced without a quoting predicate than with one (40.5%). QD initials are more often introduced by a quoting predicate (85.5%), and so are QD continuing utterances (57.4%). Thus, there is a general pattern whereby quote sequences are initiated with a quoting predicate but continuing utterances may or may not be.

Table 3: Quoting predicates in the three types of quotation sequences

Quotation Sequence Type	Quoting predicate present		Total
	N	%	N
Quote Island	237	79.5	298
Quoted Monologue initial	46	78.0	59
Quoted Monologue continuing	32	40.5	79
Quoted Dialogue initial	65	85.5	76
Quoted Dialogue continuing	105	57.4	183
Total	485	69.8	695

The distribution of active multimodal articulators per quote sequence is given in Table 4. Considering the overall use of multimodal articulators in the dataset (Total): 55.3% of quoted utterances are co-produced with character intonation, 47.6% with character facial expression, 20.4% with manual character-viewpoint (CVPT) gestures and 71.7% with the meaningful use of gaze. Although there are differences in the use of each articulator, such as the lower frequency of character-viewpoint gestures and higher frequency of the use of gaze, each multimodal articulator is co-produced with quoted utterances.

Table 4: Multimodal articulator use in the three types of quote sequences (number and proportions of quotes coded "active")

Quote Sequence Type	Character intonation		Character facial expression		Manual CVPT gesture		Meaningful use of gaze	
	N	%	N	%	N	%	N	%
Quote Island	168	56.4	133	44.6	65	21.8	220	73.8
Quoted Monologue initial	29	49.2	30	50.8	11	18.6	41	69.5
Quoted Monologue continuing	40	50.6	41	51.9	19	24.1	50	63.3
Quoted Dialogue initial	38	50.0	39	51.3	20	26.3	53	69.7
Quoted Dialogue continuing	109	59.6	88	48.1	27	14.8	134	73.3
Total	384	55.3	331	47.6	142	20.4	498	71.7

Turning to the multimodal articulation pattern of each quote sequence in Table 4, we see both similarities and differences compared to the pattern of the entire dataset. There is a general repetition of the same pattern: fewer manual character-viewpoint gestures and more meaningful uses of gaze with the other articulators falling somewhere in between. However, there appear to be differences in the frequency of some articulators in each quote sequence. For example, character intonation is the most common with QIs (56.4%) and non-initial QDs (59.6%), and the least common with initial QMs (49.2%). Character facial expression is the most common with non-initial QMs (51.9%) and initial QDs (51.3%), and the least common with QIs (44.6%). Manual character-viewpoint gesture is relatively uncommon, but is more common with initial QDs (26.1%) and less common with non-initial QDs (14.8%). And gaze is meaningfully used most often with QIs (73.8%) and non-initial QDs (73.3%), and less often with non-initial QMs (63.3%). In other words, it appears that there are different multimodal production strategies not only for each quote sequence, but also initial and non-initial utterances within those sequences.

Finally, we consider the number of articulators (0-4) accompanying each utterance in the dataset. This is presented in Table 5. While some quotes are produced with 0 active articulators (49 in the entire dataset), the majority are co-produced with 1 or more active articulators. Looking at the entire dataset, the mean number of articulators is 1.95 (sd 1.06), indicating that, on average, about two active multimodal articulators are co-produced with quoted utterances. There is some variation across the quote sequences: QIs and QDs have a slightly higher mean (1.97) while QMs have a slightly lower mean (1.88 for initial QM utterances and 1.90 for non-initial QM utterances).

Table 5: Bodily indicators of character perspective in initial and non-initial utterances in the three types of quote sequences (means and standard deviations for the occurrences of all four indicators)

Quote Sequence Type	Number of Articulators	
	Mean	Std Dev
Quote Island	1.97	1.00
Quoted Monologue initial	1.88	1.12
Quoted Monologue continuing	1.90	1.06
Quoted Dialogue initial	1.97	1.18
Quoted Dialogue continuing	1.96	1.11
Total	1.95	1.06

Considered together, these data suggest that the use of multimodal articulators varies depending on the quote sequence in which an utterance appears and the type of quote (direct speech or fictive interaction) which is produced. There also appears to be an indication of a difference in production strategies accompanying initial and non-initial utterances in multi-utterance sequences.

4.2 Modeling quote sequences on the basis of multimodal production

In the previous section, we saw that the use of multimodal articulators varied both by quote sequence and quote position (initial vs. continuing). One question we can ask is whether, based on the presence or absence of certain multimodal articulators, we are able to model the type of quote sequence in which an utterance occurs or whether the utterance is initial or non-initial. To answer these questions, we used generalized mixed linear regression (GLMER) modelling which was implemented in R 3.2.1 (R Core Team 2014) via the the glmer function in the lme4 package (Bates et al. 2014). We tested the fit of all final models using the somers2 function in the Hmisc package (Harrell 2014) which returns the index of concordance C. Logistic regression is common in other branches of linguistics (e.g. Baayen 2008, Shih et al. 2015) but not yet in gesture or sign language studies; we therefore provide an overview of the technique.

By using logistic regression, we model the probability of observing a certain outcome based on the features we enter into the model. In the final model, these features are called predictors, as their presence (or absence) has a statistically significant probability of predicting the desired outcome. This probability is given in terms of logits, the logarithm of the odds, which means that interpretation is also based on the logit scale, i.e. an estimate of 0 indicates that there is a 50% chance of observing a given outcome, an estimate > 0 indicates higher than 50%, and < 0 less than 50% chance.

To fit each model, we used a stepwise exploratory procedure whereby predictors which accounted for the least variance were eliminated one at a time. Akaike's information criterion (AIC) was used to compare models as a lower AIC indicates a better fit of the model, both in terms of the number of predictors included and the overall complexity of the model, e.g. as introduced by random intercepts and slopes (Akaike 1974). In general, an AIC difference of two is used as the minimum criteria for model selection, and indicates that the model with the lower AIC is 2.7 times more likely to provide a precise model of the data. When comparing models, the winning model should therefore have an AIC which is at least two points lower than the competing model. Finally, we measured the robustness of final models with C, the index of concordance. C indicates the amount of variance in the data which the model accounts for, and ranges from 0 to 1. Higher values indicate better performance of the model. Values of C > 0.5 indicate better than chance performance, and C > 0.8 is generally considered to be 'good' as it indicates that 80% of the variance in the data is accounted for (see Baayen 2008: p. 205, 281).

We describe a number of final models below. Each model started with the following variables of interest: quote type (reported speech vs. fictive interaction), meaningful use of gaze, character intonation, character facial expression, character-viewpoint gestures, and speaker gender. All variables were re-coded into binary variables. Variables specified by initial models were eliminated in a stepwise fashion, resulting in the final best-fit models which are presented below.[6] As the starting point for each model, we included random intercepts for speaker and narrative, and random slopes for narrative by speaker. Including random intercepts and slopes is important for preventing Type-I errors when assessing the significance of the predictors of interest (see Barr et al. 2013 and Baayen et al. 2008 for a thorough discussion on the merits of this approach). However, the data must be able to accommodate the random structures introduced by these terms. Therefore, in each model we included the maximum random effect structure supported by the data, and assessed the fit of final models using AIC and C.

---

6 Speaker gender was not predictive in any model we ran.

4.2.1 Differences across quote sequences

The first question we address is whether we can model differences between quote sequences on the basis of multimodal actions. Because our coding scheme captures the degree to which speakers demonstrate character attributes (via intonation, facial displays, manual gesture, etc.) we can test whether the narrators in our dataset embody the characters they quote, and whether that embodiment depends on certain features of the quoted sequence in which the quote occurs – such as the number of perspectives (one for QIs and QMs or two for QDs), and the extent to which a given perspective is maintained (one quoted utterance for QIs and multiple quoted utterances from one character for QMs). Both these tests yield a positive result.

We first model the number of quoted perspectives (single for QI and QM; two for QD) on the basis of multimodal actions. Because QDs are the only sequences with multiple quoted speakers, we fit a model for QD sequences (QD: 1 and not-QD: 0). The logistic regression shows that multiple speakers can indeed be distinguished from single speakers on the basis of certain multimodal articulators. The best fit GLMER model which demonstrates this is presented in Table 6.

Table 6: The best generalized mixed-effects regression model for Quoted Dialogues in the entire dataset. Only predictors for the best-fit model are shown. Negative estimates indicate lower probability. Significance is indicated as follows: *** indicates $p < 0.001$, ** $p < 0.01$, * $p < 0.05$ and . $p < 0.1$.

Predictor					
Random Effects	**Variance**				
Narrative (Intercept)	1.6436				
Speaker (Intercept)	0.8989				
Fixed Effects	**Estimate**	**Standard Error**	**z-value**	**p-value**	**Significance**
Intercept	-0.2294	0.3935	-0.583	0.5600	
Quote type: Quoted speech (0) vs. fictive interaction (1)	-2.0178	0.2804	-7.196	0.0000	***
Character facial expression: Absent (0) vs. present (1)	-0.4805	0.4232	-1.135	0.2562	
Gaze change: Absent (0) vs. present (1)	-0.6897	0.3367	-2.048	0.0405	*
Character facial expression × Gaze change	1.0975	0.4847	2.264	0.0235	*

This model has a fit of C = 0.89, which is indicative of a well-performing model. The model shows a main effect of quote type: a fictive interaction utterance is less likely to be a QD than a reported speech utterance (β = -2.0178, z = -7.196, p < 0.001). There is also a main effect of gaze change: QDs are less likely to be accompanied by gaze change (β = -0.6897, z = -2.048 , p < 0.05). Finally, there is a significant interaction between character facial expression and gaze (β = 1.0975, z = 2.264, p < 0.001). When both articulators occur, a QD is slightly more likely compared to when either articulator occurs on its own (β = -0.0727 vs. β = -0.4805 when Character facial expression is present and β = -0.6897 when only Gaze change is present).

We can also model the number of quoted speakers when only looking at multi-utterance sequences (QMs and QDs, i.e. excluding QIs). For this, we use a smaller dataset (270 utterances) composed of only the first and second position utterances in QM and QD sequences, and only QD utterances which follow an A B... pattern. We fit a model for QD sequences. The analysis shows that multiple perspectives can indeed be distinguished from single perspectives on the basis of multimodal information. The best fit GLMER model which demonstrates this is presented in Table 7.

Table 7: The best generalized mixed-effects regression model for Quoted Dialogue sequences in the smaller dataset. Only predictors for the best-fit model are shown. Negative estimates indicate lower probability. Significance is indicated as follows: *** indicates p < 0.001, ** p < 0.01, * p < 0.05 and . p < 0.1.

Predictor					
Random Effects	**Variance**				
Speaker (Intercept)	0.8495				
Fixed Effects	**Estimate**	**Standard Error**	**z-value**	**p-value**	**Significance**
Intercept	0.8531	0.3232	2.640	0.0083	**
Quote type: Quoted speech (0) vs. fictive interaction (1)	-2.2876	0.3627	-6.307	0.0000	***
Character facial expression: Absent (0) vs. present (1)	0.7429	0.3386	2.194	0.0282	*

This model has a fit of C = 0.83, which is indicative of a well-performing model. The model shows a main effect of quote type: a fictive interaction utterance is less likely to be a QD than a direct speech utterance (β = -2.2876, z = -6.387, p < 0.001). There is also a main effect of character facial expression: QDs are more likely to be accompanied by character facial expression (β = 0.7429, z = 2.194, p < 0.05).

As the models in Tables 6 and 7 show, the difference between multiple and single speakers is the use of fictive interaction (more likely with single speakers), the use of meaningful gaze change (more likely with single speakers in the entire dataset), the use of character facial expression (more likely with multiple speakers in the smaller dataset) and the interaction between character facial expression and gaze (their co-occurrence is more likely with multiple speakers in the larger dataset). Thus, it is possible to model the number of quoted speakers on the basis of multimodal activity: there appears to be more multimodal activity accompanying quotes from single quoted speakers (QI or QM) than multiple ones (QD).

Next, we investigate differences in quotes of single speakers (not-QDs in the models presented above). To do this, we model whether single or multiple utterances from a speaker were quoted. Because QIs and QMs both involve a single speaker, this is equivalent to asking if we can generate different models for QI and QM sequences in the entire dataset. As Tables 8 and 9 show, the answer appears to be yes: there are differences in the multimodal production of QI and QM sequences, with QI sequences garnering more predictors than QM sequences.

Table 8: The best generalized mixed-effects regression model for Quote Islands. Only predictors for the best-fit model are shown. Negative estimates indicate lower probability. Significance is indicated as follows: *** indicates p < 0.001, ** p < 0.01, * p < 0.05 and . p < 0.1.

Predictor					
Random Effects	**Variance**				
Narrative (Intercept)	0.8510				
Speaker (Intercept)	0.5994				
Fixed Effects	**Estimate**	**Standard Error**	**z-value**	**p-value**	**Significance**
Intercept	-1.2602	0.3700	-3.406	0.0007	***
Quote type: Quoted speech (0) vs. fictive interaction (1)	0.7274	0.2111	3.446	0.0006	***
Character facial expression: Absent (0) vs. present (1)	0.5512	0.3652	1.509	0.1313	
Gaze change: Absent (0) vs. present (1)	0.8217	0.2900	2.834	0.0046	**
Quoting predicate: Verb (1) vs. bare (0)	0.8111	0.2320	3.496	0.0005	***
Character facial expression × Gaze change	-0.8769	0.4160	-2.108	0.0350	*

The best GLMER model for QIs is presented in Table 8. It has a fit of C = 0.83, which is indicative of a well-performing model. The model shows a main effect of quote type: a fictive interaction utterance is more likely to be a QI than a direct speech utterance (β = 0.7274, z = 3.446, p < 0.001). In addition, there is a main effect of gaze change: QIs are likely to be produced with a meaningful use of gaze (β = 0.8217, z = 2.834, p < 0.01). There is also a main effect of quoting predicate: QIs are more likely to be accompanied by a quoting verb than a bare quote (β = 0.8111, z = 3.496, p < 0.001). Finally, there is a significant interaction between character facial expression and gaze (β = -0.8769, z = -2.108, p < 0.05). When both occur, a QI is slightly less likely compared to when either occurs alone (β = 0.496 vs. β = 0.5512, when only Character facial expression is present and β = 0.8217, when only Gaze change is present).

Table 9: The best generalized mixed-effects regression model for Quoted Monologues. Only predictors for the best-fit model are shown. Negative estimates indicate lower probability. Significance is indicated as follows: *** indicates p < 0.001, ** p < 0.01, * p < 0.05 and . p < 0.1.

Predictor					
Random Effects	**Variance**				
Narrative (Intercept)	0.9298				
Speaker (Intercept)	0.1796				
Fixed Effects	**Estimate**	**Standard Error**	**z-value**	**p-value**	**Significance**
Intercept	-1.5194	0.3478	-4.369	0.0000	***
Quote type: Quoted speech (0) vs. fictive interaction (1)	1.0537	0.2390	4.408	0.0000	***
Character intonation: Absent (0) vs. present (1)	-0.4172	0.2277	-1.832	0.0669	.
Quoting predicate: Verb (1) vs. bare (0)	-0.9302	0.2522	-3.688	0.0002	***

The best GLMER model for QMs is presented in Table 9. It has a fit of C = 0.82, which is good. The model shows a main effect of quote type: a fictive interaction utterance is more likely to be a QM than a direct speech utterance (β = 1.0537, z = 4.408, p < 0.001). In addition, there is a marginal effect of character intonation: QMs are less likely to be co-produced with character intonation (β = -0.4172, z = -1.832, p < 0.1). There is also a main effect of quoting predicate: QM's are more likely to be accompanied by a quoting verb than a bare quote (β = -0.9302, z = -3.688, p < 0.001).

Taken together, these results indicate that it is possible to model both the number of quoted speakers and the kind of quote sequence on the basis of multimodal actions: there is more evidence of multimodal activity which demonstrates character perspective in brief single speaker quotes (QIs) than extended single speaker quotes (QMs) or quotes which are part of multiple speaker sequences (QDs) .

4.2.2 Differences within quote sequences

Another question we can ask concerns the internal organization of quote sequences, and whether it is possible to model initial vs. non-initial utterances. There are two ways of asking and answering this question: (1) Is there a difference between initial or continuing utterances in the whole dataset? (2) Looking only at multi-utterance sequences (i.e. QMs and QDs), is there a difference between the initial utterance in the sequence or the second utterance in the sequence? We will consider these questions in turn.

4.2.2.1 Initial *vs.* continuing utterances in the whole dataset

For this analysis, we ask if it is possible to model initial utterances (QIs, QM initial and QD initial) in the whole dataset.[7] As Table 10 shows, this is possible – but, given the results presented earlier, we suspect this is driven by the large number of QI utterances: 179 QI, 59 QM initial and 76 QD initial utterances.

---

7 We do not report models for second position utterances in the entire dataset or initial utterances in the smaller dataset as the fit of the models (C = 0.63 and 0.61, respectively) was indicative of poor-performing models.

Table 10: The best generalized mixed-effects regression model for initial utterances in any quote sequence. Only predictors for the best-fit model are shown. Negative estimates indicate lower probability. Significance is indicated as follows: *** indicates p < 0.001, ** p < 0.01, * p < 0.05 and . p < 0.1.

Predictor					
Random Effects	**Variance**				
Narrative (Intercept)	0.1771				
Speaker (Intercept)	0.3588				
Fixed Effects	**Estimate**	**Standard Error**	**z-value**	**p-value**	**Significance**
Intercept	-0.7127	0.3152	-2.262	0.0237	*
Quote type: Quoted speech (0) vs. fictive interaction (1)	0.5577	0.2080	2.681	0.0073	**
Character facial expression: Absent (0) vs. present (1)	0.4220	0.3526	1.197	0.2314	
Gaze change: Absent (0) vs. present (1)	0.6131	0.2831	2.166	0.0303	*
Quoting predicate: Verb (1) vs. bare (0)	1.4924	0.2150	6.941	0.0000	***
Character facial expression × Gaze change	-0.7681	0.4066	-1.889	0.0589	.

The best GLMER model for initial utterances is presented in Table 13. It has a fit of C = 0.79, which is indicative of a fairly well-performing model. The model shows a main effect of quote type: a fictive interaction utterance is more likely to be an initial utterance ($\beta = 0.5577$, z = 2.681, p < 0.01). There is also a main effect of the meaningful use of gaze, which is more likely to accompany initial utterances ($\beta = 0.6131$, z = 2.166, p < 0.05). In addition, there is a main effect of quoting predicate; initial utterances are more likely to be introduced by a quoting verb than a bare quote ($\beta = 1.4924$, z = 6.941, p < 0.001). Finally, there is a marginal interaction between character facial expression and gaze ($\beta = -0.7681$, z = -1.889, p < 0.1). When both occur, an initial utterance is slightly less likely to occur compared to when the articulators occur alone ($\beta = 0.267$ vs. $\beta = 0.4220$, when only character facial expression is present and $\beta = 0.6131$, when only gaze is present).

These results suggest two things: First, initial utterances may be accompanied by more multimodal articulators – but this result needs further investigation, since it is possible that the large number of QIs is driving the model. Second, utterances in the continuing position of quote sequences appear not to be accompanied by much multimodal activity – but this result also needs further investigation, since it is possible that within each QD and QM a different articulatory pattern might emerge.

4.2.2.2 Initial *vs.* second utterances in QMs and QDs

To answer the question whether it is possible to model the initial vs. continuing utterances of multi-utterance sequences on the basis of multimodal actions, we use a restricted dataset which is composed only of QM and QD sequences, and only utterances which occur in the first or second position of those sequences. This restricted dataset is composed of 270 utterances, and only QD sequences which follow the A B... pattern. We ask if it is possible to predict the initial utterance in each type of sequence.[8]

Tables 11 presents the best fit model for initial QM utterances, and Tables 12 presents the best fit model for initial QD utterances. While these results indicate a complementary pattern whereby QMs are more likely to be instances of fictive interaction with a quoting predicate in first position and a bare quote in second position, and QDs are more likely to be instances of direct speech (but not fictive interaction) with a quoting predicate in first position and a bare quote in second position, there are no multimodal predictors in either model.

---

8 Models for first and second position QM and QD utterances in the smaller dataset are identical, but with opposite signs on the estimates. This is reported in the paper package.

Table 11: The best generalized mixed-effects regression model for initial utterances in Quoted Monologue sequences. Only predictors for the best-fit model are shown. Negative estimates indicate lower probability. Significance is indicated as follows: *** indicates p < 0.001, ** p < 0.01, * p < 0.05 and . p < 0.1.

Predictor					
Random Effects	**Variance**				
Speaker (Intercept)	0.02958				
Fixed Effects	**Estimate**	**Standard Error**	**z-value**	**p-value**	**Significance**
Intercept	-2.3053	0.4100	-5.622	0.0000	***
Quote type: Quoted speech (0) vs. fictive interaction (1)	1.1469	0.3217	3.565	0.0036	***
Quoting predicate: Verb (1) vs. bare (0)	0.7920	0.3855	2.055	0.0399	*

The best GLMER model for initial QM utterances is presented in Table 11. It has a fit of C = 0.70, which is indicative of a fair model. The model shows a main effect of quote type: fictive interaction is more likely to appear in the first position of a QM (β = 1.1469, z = 3.565, p < 0.001). There is also a main effect of quoting predicate; first position QM utterances are more likely to be accompanied by a quoting verb (β = 0.792, z = 2.055, p < 0.05).

Table 12: The best generalized mixed-effects regression model for initial utterances in Quoted Dialogue sequences. Only predictors for the best-fit model are shown. Negative estimates indicate lower probability. Significance is indicated as follows: *** indicates p < 0.001, ** p < 0.01, * p < 0.05 and . p < 0.1.

Predictor					
Random Effects	**Variance**				
Speaker (Intercept)	0.0000				
Fixed Effects	**Estimate**	**Standard Error**	**z-value**	**p-value**	**Significance**
Intercept	-1.4205	0.3401	-4.177	0.0000	***
Quote type: Quoted speech (0) vs. fictive interaction (1)	-1.3871	0.3747	-3.701	0.0002	***
Quoting predicate: Verb (1) vs. bare (0)	1.0956	0.3671	2.985	0.0028	**

The best GLMER model for initial QD utterances is presented in Table 12. It has a fit of C = 0.73, which is indicative of a fair model. The model shows a main effect of quote type: fictive interaction is less likely to appear in the first position of a QD (β = -1.3871, z = -3.701, p < 0.001). There is also a main effect of quoting predicate; first position QD utterances are more likely to be accompanied by a quoting verb (β = 1.0956, z = 2.985, p < 0.01).

The models presented in Tables 11 and 12 are less robust than the models presented above (values of C around 0.7). This might reflect the size of the dataset, and the small number of quotes of each type investigated in these analyses (59 QM initial and continuing utterances; 76 QD initial and continuing utterances). With a larger dataset, models with the multimodal predictors we identified might emerge. Alternatively, the models could indicate that, in fact, the multimodal features we identified are not sufficient for modeling differences in the production of utterances within quote sequences. This does not necessarily mean there is no multimodal difference – it might be that the size of the speaker's gesture space, the degree to which their articulators are tensed or lax, or other linguistic or multimodal features not considered might play a role.

4.2.3 Summary of regression results

In summary, an overview of all final models is presented in Table 13. For each model presented, this table shows which features are predictive and whether the estimate for that feature is positive or negative, i.e. whether the feature is more or less likely to occur. Note that the models are presented by dataset, not in presentation order.

Table 13: Overview of all regression results. Significance is indicated as follows: +++/--- indicates $p < 0.001$, ++/-- $p < 0.01$, +/- $p < 0.05$ and (+)/(-) $p < 0.1$, with + or - marking positive or negative estimates. C indicates the fit of each model.

	Entire dataset				Small dataset		
	Initial	QI	QM	QD	Multiple Speakers	QMi	QDi
Quote type	++	+++	+++	---	---	+++	---
Character facial expression					+		
Gaze change	+	++		-			
Character facial expression x Gaze	(+)	-		+			
Character intonation			(-)				
Quoting Predicate	+++	+++	---			+	++
C	0.79	0.83	0.82	0.89	0.83	0.7	0.73

As our results demonstrate, the active use of multimodal articulators (character intonation, character facial expression, manual character-viewpoint gestures and the meaningful use of gaze as we define it) appears to play a role in the production of multimodal quotes; all are co-produced with quotes some of the time, but only some are predictive of quote sequence or quote position. In line with previous qualitative research, character intonation, character facial expression and the meaningful use of gaze are found to be predictive while manual character viewpoint gestures are not. In addition, while these multimodal articulators are predictive of initial utterances in the entire dataset, they fail to be significant in predicting initial QM or QD utterances in the smaller dataset. However, they do appear to play a role with respect to the type of quote sequence, for which character facial expression, the meaningful use of gaze, and their interaction were predictive in both datasets.

5 Discussion and conclusions

In this study, we investigated the multimodal production of quoted utterances in three types of quote sequences: Quote Islands (QIs), Quoted Monologues (QMs) and Quoted Dialogues (QDs). Going beyond previous research which focused on qualitative analyses of the production of multimodal direct speech utterances (often about single quoted utterances – e.g. Fox & Robles 2010; Sidnell 2006), we provided a comprehensive quantitative comparison of the production of multimodal quote sequences, which account for possible differences in the number of quoted speakers and number of quoted utterances by those speakers. While multimodal articulators and linguistic features were found to be predictive of quote sequences, speaker gender was not. This suggests a multimodal differentiation of quotation contexts which is true of a community of speakers rather than one which stems from the individual characteristics indicated here. We discuss these community-wide practices below.

5.1 Number of quoted speakers

We investigated the extent to which the number of quoted speakers affects the production of multimodal quotes. Previous work has demonstrated the coordination of multiple multimodal articulators during direct speech utterance production (Sidnell 2006; Park 2009), but has largely been focused on single utterances by single quoted speakers. Our study extends this body of work by considering the multimodal production of Quoted Monologues and Quoted Dialogues alongside single quotes, or Quote Islands. As our analysis demonstrates, different multimodal production strategies were used for multiple quoted speakers (QDs vs. QMs + QIs). Moreover, different multimodal articulators were used when quoting one speaker one time (QIs) or multiple times (QMs). These differences were demonstrated by the best-fit models produced for each type of quote sequence (Section 4.2.1). As expected, QD utterances attracted more multimodal indicators of character perspective, and single utterance quotations (QIs) were treated differently from multi-utterance sequences (QMs + QDs). The models supporting these statements were the most robust models we produced

– and clearly invite further research into the means by which multiple articulators co-occur and express character viewpoint.

5.2 Transition into quote sequences

As previous research has identified multimodal articulation at the start of quoted utterances (Sidnell 2006), we also investigated the extent to which the position of the quoted utterance affects multimodal production. To do this, we focused on the multimodal production of initial utterances (QI, QM initial and QD initial utterances) compared to continuing utterances (QM non-initial and QD non-initial utterances), as well as the multimodal production of initial compared to second utterances within multi-utterance sequences (QM and QD utterances). As the models in Section 4.2.2 demonstrate, there were multimodal indicators of perspective shift for initial utterances in the entire dataset – but this finding was driven by the inclusion of QI utterances. When investigating initial QM and initial QD utterances in the smaller dataset – composed only of QM and QD sequences – this multimodal differentiation was not present. Instead, the best-fit models for initial vs. non-initial utterance comparisons identified quote type (direct speech vs. fictive interaction) and quoting predicate (any predicate vs. bare quotes) as the best models of initial utterances. As previous work is quite consistent about demonstrating multimodal utterance production at the start of quotes, we would have expected initial utterances to garner more (or different) multimodal articulators than continuing utterances and so find this "null" result (null in the sense that multimodal articulators were not predictive) puzzling. It might be the case that the expression of character viewpoint is not the best measure of these changes – another feature such as the use of gestural space or the tension of multimodal articulators might be. It might also be the case that a dataset with more QM and QD utterances would be better able to identify a difference.

One important finding which all of our analyses touch on concerns the simultaneous use of multimodal articulators. Our research identifies some of the bodily articulators which are relevant for distinguishing perspective or embodying characters when quoting in casual oral narration. Whereas previous research (e.g. Park 2009) provided a qualitative account of the contribution of multiple articulators to the production of direct speech utterances, we were able to quantify this activity and demonstrate that the co-occurrence of multimodal articulators has some predictive properties. Like Earis & Cormier (2013), who investigated the expression of character viewpoint in narrative settings, we found the use of character intonation and character facial expression was more frequent than that of manual character viewpoint gestures. Similarly, we found that character facial expression and a meaningful use of gaze co-occurred. In other words, we were able to identify specific bodily articulators which play a role in multimodal quotation, and how their use varies by environment (type of quotation sequence). Our results indicated a mean co-occurrence of about two articulators per quote, but further investigation is needed to more fully understand how linguistic and paralinguistic indicators and bodily articulators jointly cue shifts to character viewpoint and perspective shifts in general.

Finally, researchers interested in the perspective-taking abilities of speaking populations generally look at overall narrative production strategies and how, for example, certain event or linguistic properties constrain manual gesture production within them (e.g. McNeill 1992; Parrill 2010, among many others). However, as our investigation has demonstrated, multimodal quoted utterances are also important for documenting the multimodal perspective-taking abilities of speakers. While there are some indications that direct speech is considered to be more lively and vivid than indirect speech (e.g. Li 1986; Groenewold et al. 2014) and that this difference might be linked to certain prosodic or paralinguistic features (Yao et al. 2012), no links between these findings and multimodal articulation or character embodiment have yet been identified. Our investigation suggests that one way direct speech quotation might accomplish this is through the use of multimodal bodily actions which demonstrate certain aspects of the quoted character.

References

Akaike, Hirotugu. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, pp. 716–723.

Baayen, R. Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

Baayen, R. Harald, Douglas J. Davidson, & Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59* (4), pp. 390–412.

Bakhtin, Micheal M. 1981. *The Dialogic Imagination*. Austin: University of Texas Press.

Banfield, Ann. 1982. *Unspeakable sentences: Narration and representation in the language of fiction*. Routledge.

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68* (3), pp. 255-278.

Bates, Douglas, Martin Maechler, Ben Bolker & Steven Walker. 2014. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. Retrieved from: http://CRAN.R-project.org/package=lme4.

Berko-Gleason, Jean, Harold Goodglass, Loraine Obler, Eugene Green, Mary R. Hyde, & Sandra Weintraub. 1980. Narrative strategies of aphasic and normal-speaking subjects. *Journal of Speech, Language, and Hearing Research*, *23*(2), pp. 370-382.

Buchstaller, Isabelle. 2013. *Quotatives: new trends and sociolinguistic implications*. Hoboken, N.J.: John Wiley & Sons.

Buchstaller, Isabelle, & Alexandra D'Arcy. 2009. Localized globalization: A multi-local, multivariate investigation of quotative *be like*. *Journal of Sociolinguistics*, *13*(3), pp. 291-331.

Chovil, Nicole. 1991. Discourse-oriented facial displays in conversation. *Research on Language & Social Interaction, 25* (1-4), pp. 163-194.

Clark, Herbert H., & Richard J. Gerrig. 1990. Quotations as demonstrations. *Language, 66* (4), pp. 764-805.

Clark, Herbert H., and Mija M. Van Der Wege. 2001. Imagination in discourse. In Deborah Shriffrin, Deborah Tannen & Heidi E. Hamilton (Eds.),*The handbook of discourse analysis*, pp. 772-786. Oxford: Blackwell Publishing.

Cormier, Kearsy, David Quinto-Pozos, Zed Sevcikova, & Adam Schembri. 2012. Lexicalisation and de-lexicalisation processes in sign languages: Comparing depicting constructions and viewpoint gestures. *Language and Communication, 32*(4), pp. 329-48.

Cormier, Kearsy, Sandra Smith, & Martine Zwets. 2013. Framing constructed action in British Sign Language narratives. *Journal of Pragmatics*, *55*, pp. 119-139.

Couper-Kuhlen, Elizabeth. 1999. Coherent voicing: On prosody in conversational reported speech. *Pragmatics and Beyond New Series*, pp. 11-34.

Couper-Kuhlen, Elizabeth. 2007. Assessing and accounting. In Elizabeth Holt & Rebecca Clift (Eds.), *Reporting Talk. Reported Speech in Interaction*, pp.81-119. Cambridge: Cambridge University Press.

Dancygier, Barbara & Eve Sweetser (Eds.). 2012. *Viewpoint in language: A multimodal perspective.* Cambridge University Press.

Earis, Helen & Kearsy Cormier. 2013. Point of view in British Sign Language and spoken English narrative discourse: the example of "The Tortoise and the Hare". *Language and Cognition*, *5*(1), pp. 313-343.

Eerland, Anita, Jan A. A. Engelen, & Rolf A. Zwaan. 2013. The influence of direct and indirect speech on mental representations. *PLoS One*, *8*, e65480.

Engberg-Pedersen, Elisabeth. 1993. *Space in Danish Sign Language: The semantics and morphosyntax of the use of space in a visual language*. SIGNUM-Press.

Fox, Barbara A., & Jessica Robles. 2010. *It's like mmm*: Enactments with it's like. *Discourse Studies, 12,* pp. 715-738.

Gnisci, Augusto, Fridanna Maricchiolo, & Marino Bonaiuto. 2014. Reliability and validity of coding systems for bodily forms of communication. In Cornelia Müller, Alan Cienki, Ellen Fricke et al. (Eds.), *Body Language Communication,* pp. 879-892. Berlin: Mouton de Gruyter.

Goodwin, Charles. 2007. Interactive footing. In Elisabeth Holt & Rebecca Clift (Eds.). *Reporting Talk. Reported Speech in Interaction*, pp.16-46. Cambridge: Cambridge University Press.

Groenewold, Rimke, Roelien Bastiaanse, Lynsey Nickels & Mike Huiskes. 2014. Perceived liveliness and speech comprehensibility in aphasia: the effects of direct speech in auditory narratives. *International Journal of Language & Communication Disorders, 49* (4), pp. 486-497.

Guerini, Federica. 2015. Being a former Second World War partisan: Reported speech and the expression of local identity. *Open Linguistics*, *1*(1), pp. 191-210.

Günthner, Susanne. 1999. Polyphony and the 'layering of voices' in reported dialogues: An analysis of the use of prosodic devices in everyday reported speech. *Journal of Pragmatics, 31*(5)*,* pp. 685-708.

Harrell, Frank E., Jr. 2014. Hmisc package version 3.14-6. http://cran.r-project.org/web/packages/Hmisc/index.html

Hengst, Julie A., Simone R. Frame, Tiffany Neuman-Stritzel, & Rachel Gannaway. 2005. Using others' words: Conversational uses of reported speech by individuals with aphasia and their communicative partners. *Journal of Speech, Language and Hearing Research, 48*, pp. 137-156.

Holler, Judith, & Katie Wilkin. 2009. Communicating common ground: How mutually shared knowledge influences speech and gesture in a narrative task. *Language and Cognitive Processes*, *24*(2), pp. 267-289.

Janzen, Terry. 2012. Two ways of conceptualizing space. In Barbara Dancygier & Eve Sweetser (Eds.), *Viewpoint in Language: A Multimodal Perspective*, pp.156-175. Cambridge: Cambridge University Press.

Kendon, Adam. 2004. *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Labov, William. 1972. *Language in the city: Studies in the Black English vernacular*. Philadelphia: University of Pennsylvania Press.

Lehrer, Adrienne. 1989. Remembering and representing prose: Quoted speech as a data source. *Discourse Processes, 12*(1), pp. 105–125.

Levy, Elena T., & David McNeill. 1992. Speech, gesture, and discourse. *Discourse Processes*, *15*(3), 277-301.

Li, Charles. N. 1986. Direct and Indirect Speech: a functional study. In Florian Coulmas (Ed.). *Direct and Indirect Speech*, pp.29-45. Berlin: Mouton de Gruyter.

Liddell, Scott. K., & Melanie Metzger. 1998. Gesture in sign language discourse. *Journal of Pragmatics, 30*(6), pp. 657-697.

Mayes, Patricia. 1990. Quotation in spoken English. *Studies in Language, 14*, pp. 325–363.

McClave, Evelyn Z. 2000. Linguistic functions of head movements in the context of speech. *Journal of Pragmatics*, *32*(7), pp. 855-878.

McNeill, David. 1992. *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

Metzger, Melanie 1995. Constructed dialogue and constructed action in American Sign Language. In Ceil Lucas (Ed.), *Sociolinguistics in Deaf Communities*, pp.255-271. Washington, DC: Gallaudet University Press.

Özyürek, Asli. 2002. Do speakers design their cospeech gestures for their addressees? The effects of addressee location on representational gestures. *Journal of Memory and Language*, *46*(4), pp. 688-704.

Padden, Carol A. 1986. Verbs and role shifting in American Sign Language. In Carol Padden (Ed.), *Proceedings of the Fourth National Symposium on Sign Language Research and Teaching Sign*, pp.44-57. Silver Spring, MD: NAD.

Park, Yujong. 2009. Interaction between grammar and multimodal resources: quoting different characters in Korean multiparty conversation. *Discourse Studies, 11*(1), pp. 79-104.

Parrill, Fey. 2010. The hands are part of the package: Gesture, common ground, and information packaging. In John Newman & Sally Rice (Eds.), *Empirical and Experimental Methods in Cognitive/Functional Research*, pp. 285-302. Stanford: CSLI Publications.

Parrill, Fey. 2012. Interactions between discourse status and viewpoint in co-speech gesture. In Barbara Dancygier & Eve Sweetser (Eds.), *Viewpoint in language: A multimodal perspective*, pp.97-112. Cambridge: Cambridge University Press.

Pascual, Esther. 2014. *Fictive interaction: The conversation frame in thought, language, and discourse*. Amsterdam: John Benjamins Publishing Company.

Perniss, Pamela, & Asli Özyürek. 2015. Visible cohesion: A comparison of reference tracking in sign, speech, and co-speech gesture. *Topics in Cognitive Science*, *7*(1), pp. 36-60.

Perniss, Pamela, Asli Özyürek, & Gary Morgan. 2015. The influence of the visual modality on language structure and conventionalization. *Topics in Cognitive Science, 7*(1), 2-11.

R Core Team. 2014. R: A Language And Environment For Statistical Computing. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/.

Redeker, Gisela. 1991. Linguistic markers of discourse structure. *Linguistics*, *29*(6), pp. 1139-1172.

Sakita, Tomoko I. 2002. *Reporting Discourse, Tense, and Cognition*. Oxford: Elsevier.

Sams, Jessie. 2010. Quoting the unspoken: An analysis of quotations in spoken discourse. *Journal of Pragmatics, 42*(11), pp. 3147-3160.

Shih, Stephanie, Jason Grafmiller, Richard Futrell, & Joan Bresnan. 2015. Rhythm's role in genitive construction choice in spoken English. In Ralf Vogel & Ruben Vijver (Eds.), *Rhythm in cognition and grammar,* pp. 207-234. Berlin: De Gruyter Mouton.

Sidnell, Jack. 2006. Coordinating gesture, talk, and gaze in reenactments. *Research on Language and Social Interaction,* *39*(4), pp. 377-409.

Simmons, Katie, & Amanda LeCouteur. 2011. 'Hypothetical active-voicing': Therapists 'modelling' of clients' future conversations in CBT interactions. *Journal of Pragmatics, 43*(13), pp. 3177-3192.

Stec, Kashmiri. 2012. Meaningful shifts: A review of viewpoint markers in co-speech gesture and sign language. *Gesture,* *12*(3), pp. 327-360.

Stec, Kashmiri, Mike Huiskes, Alan Cienki, & Gisela Redeker. (Under review). Annotating bodily indicators of perspective shifts in conversational narratives.

Stec, Kashmiri, Mike Huiskes, & Martijn Wieling. (Under review). Multimodal character viewpoint in quoted dialogue sequences

Stelma, Juurd H., & Lynne J. Cameron. 2007. Intonation units in spoken interaction: Developing transcription skills. *Text & Talk-An Interdisciplinary Journal of Language, Discourse Communication Studies*, *27*(3), pp. 361-393.

Streeck, Jurgen, & Mark L. Knapp. 1992. The interaction of visual and verbal features in human communication. In Fernando Potayos (Ed.), *Advances in non-verbal communication: Sociocultural, clinical, esthetic and literary perspectives*, pp.3-23. Amsterdam: John Benjamins.

Sweetser, Eve, & Kashmiri Stec. (In press). Maintaining multiple viewpoints with gaze.

Tannen, Deborah. 1989. *Talking voices: Repetition, dialogue, and imagery in conversational discourse*. Cambridge University Press.

Thompson, Sandra A., & Ryoko Suzuki. 2014. Reenactments in conversation: Gaze and recipiency. *Discourse Studies*, *16*(6), pp. 1-31.

Turner, Mark. 1998. *The literary mind: The origins of thought and language*. Oxford: Oxford University Press.

Uebersax, John S. 1987. Diversity of decision-making models and the measurement of interrater agreement. *Psychological Bulletin*, *101(1)*, pp. 140-146.

Vigliocco, Gabriella, Pamela Perniss, & David Vinson. 2014. Language as a multimodal phenomenon: Implications for language learning, processing and evolution. *Philosophical Transaction of the Royal Society B*, 369 (20130292). DOI: 10.1098/rstb.2013.0292

Wittenburg, Peter, Hennie Brugman, Albert Russel, Alex Klassmann, & Han Sloetjes. 2006. ELAN: a Professional Framework for Multimodality Research. In: *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*. Retrieved from http://pubman.mpdl.mpg.de/pubman/item/escidoc:60436:2/component/escidoc:60437/LREC%202006_Elan_Wittenburg.pdf

Yao, Bo, Pascal Belin, & Christoph Scheepers. 2012. Brain "talks over" boring quotes: Top-down activation of voice-selective areas while listening to monotonous direct speech quotations. *NeuroImage, 60*(3), pp. 1832–1842.

Zunshine, Lisa. 2006. *Why we read fiction: Theory of mind and the novel*. Columbus: Ohio State University Press.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51