Language & Communication 88 (2023) 99–110

Contents lists available at ScienceDirect

Language & Communication

journal homepage: www.elsevier.com/locate/langcom

Contextualization cues for media references in everyday conversation☆

Sylvia Sierra

Syracuse University, 115 Sims Hall, Syracuse, NY, 13244, USA

ARTICLE INFO

Article history:

Keywords:
Conversation
Contextualization cues
Play
Intertextuality
Media
Identity

ABSTRACT

While scholars have explored the importance of quoting media in accomplishing relationship and identity work in conversation, there is little work on how speakers phonetically and paralinguistically signal spoken media references specifically so that they may be recognized in the speech stream. This article demonstrates how speakers make 148 media references recognizable across 5 audio-recorded everyday conversations among friends. I identify 5 ways that these playful media references are signaled in talk: word stress and particular intonation contours, pitch register shifts, smiling and laughter, performing stylized accents, and singing. This systematic analysis of the contextualization cues used to signal media references in everyday talk contributes to understanding how speakers actively participate in intertextual processes.

1. Introduction

Scholars have explored the interactional importance of the appropriation of media in everyday talk in accomplishing relationship and identity work (e.g. Ayaß & Gerhardt 2012; Duff 2002; Spitulnik 1996; Tovares 2006, 2007). Most recently, for instance, I have examined the use of playful media references in shifting frames when speakers are faced with epistemic imbalances in talk, ultimately functioning to construct shared identities (Sierra, 2016a, b; Sierra 2021). Yet there has been scant research on how in spoken language people phonetically signal that they are referencing media in talk in order to make such references recognizable to listeners. Tetreault (2009) represents one attempt to explain how mimicking a TV host is phonetically signaled in talk, but the phonological and paralinguistic features are only briefly described, with more of a focus on lexical and grammatical features. At the same time, there is a substantial body of work on how reported speech is signaled in everyday talk (e.g., Bauman 1992; Couper-Kuhlen 1996, 1999; De Decker 2013; Fletcher 2005; Gumperz 1992; Günthner 1992, 1999; Hirschberg and Grosz 1992; Jansen et al., 2001; Jones 2016; Klewitz and Couper-Kuhlen 1999; Kotthoff 2021; Park 2009; Schwitalla 1992; Tannen, 1989/2007).

In this paper, I build on this prior work to explore how intertextual references to both written and oral, as well as printed and electronic media such as films, TV shows, audio recordings, internet memes, and songs in audio-recorded everyday conversation among friends are made recognizable in spoken speech not only via lexical repetition in the absence of a quotative (Mathis and Yule 1994), but also through specific prosodic, paralinguistic, and phonetic features. I use Gumperz's (1977, 1982) work on 'contextualization cues' and apply this concept specifically to playful media references in conversation. By describing the contextualization cues used to signal media references in talk, I show that these cues play an important role

☆ This article is a condensed and more technical version of chapter 2 of Sierra 2021; example 1 in this article is presented in chapter 4 of the same work.

E-mail address: ssierra@syr.edu.

https://doi.org/10.1016/j.langcom.2022.11.003
0271-5309/

in allowing speakers to recognize and build on intertextual media references. This paper contributes to our understanding of the fundamentals of how speakers actively engage in intertextual processes in interaction.

2. Contextualization cues for intertextuality and play

Gumperz (1977, 1982) developed the concept of a conversational 'meta-signaling system' (1977: 192) in which contextualization cues are defined as 'any aspect of the surface form of utterances which, when mapped onto message content, can be shown to be functional in the signaling of interpretative frames' (1977: 199). In other words, these cues are components of language, such as pitch, intonation, stress, rhythm, tempo, loudness, laughter, choice of code, particular lexical expressions, and nonverbal cues (such as gestures, facial expressions, and gaze) that carry meaning (see also Duranti & Goodwin 1992: 229). When combined with the semantic content of utterances, they signal to listeners how the message ought to be interpreted. Contextualization cues enable 'conversational inference': the immediate process of interpreting meaning, in which conversational participants assess interlocutors' intentions, and on which they form responses (Gumperz 1977: 191).

Further empirical studies have followed Gumperz's foundational work on contextualization cues. Auer and Di Luzio (1992) and Duranti & Goodwin (1992) provide edited volumes containing much of this work, covering contextualization across a variety of topics, including but not limited to the lexical choice in contextualizing respect (Duranti 1992), deictic reference (Hanks 1992), turn-taking (Streeck and Hartge 1992), assessments (Goodwin and Goodwin 1992), repair (Couper-Kuhlen 1992; Philips 1992), frames (Gaik 1992), narratives (Basso 1992; Bauman 1992; Selting 1992; see also Selting 1994; Uhmann 1996), and performance and genres (Bauman 1992).

Other empirical studies have examined specifically how reported speech is contextualized in interaction. While many studies have focused on the use of quotatives such as *go* or *be like* in intertextually voicing others in American English interaction (Blyth et al., 1990; Buchstaller & D'Arcy 2009; D'Arcy 2017; Romaine and Lange 1991; Tannen 1986; Yule and Mathis, 1992), Mathis and Yule (1994) bring attention to the use of "zero quotatives" where direct speech is reported with neither a reporting verb nor an attributed speaker. They find that shifts in pitch, changes in voice quality (e.g. breathy voice, creaky voice, falsetto), and loudness are all used to signal reported speech in zero quotatives. They make a connection between this practice in everyday conversation and "intonational quotation marks" which Bakhtin (1981) identifies in his analysis of literary texts.

Indeed, shifts in pitch, changes in voice quality, and loudness are also found by many other researchers to be indicative of signaling another 'voice', 'reported dialogue', 'reported speech,' or 'constructed dialogue' (Tannen, 1989/2007) in interaction. Shifts in pitch are perhaps the most widely documented contextualization cue for reported speech (e.g., Couper-Kuhlen 1996, 1999; De Decker 2013; Fletcher 2005; Gumperz 1992; Günthner 1999; Hirschberg and Grosz 1992; Jansen et al., 2001; Jones 2016; Klewitz and Couper-Kuhlen 1999; Kotthoff 2021). Changes in voice quality are also common in signaling other 'voices' (e.g. Couper-Kuhlen 1996, 1999; Gumperz 1992; Günthner 1992, 1999; Klewitz and Couper-Kuhlen 1999; Park 2009; Tannen, 1989/2007). Finally, loudness is also commonly used to signal constructed dialogue (Bauman 1992; Couper-Kuhlen 1996, 1999; Günthner 1999; Klewitz and Couper-Kuhlen 1999; Tannen, 1989/2007). In addition, researchers have also found that distinct intonation patterns can signal reported speech (Couper-Kuhlen 1996, 1999, Günthner 1999, Schwitalla 1992, Tannen, 1989/2007) and so can tempo and rhythmic pattern (Bauman 1992; Couper-Kuhlen 1996, 1999; Günthner 1999; Klewitz and Couper-Kuhlen 1999). Additionally, Park (2009) shows smile voice and laughter can be used along with a change in voice quality when quoting a non-present third-party.

Tetreault (2009) partially describes how mimicking media is phonetically contextualized. The primary contextualizing phonological feature used to mock a TV host is described as "crisp pronunciation" (Tetreault 2009: 208). More precisely described is a single use of emphatic stress on the first syllable. Overall, these prior studies highlight the phonetic work that speakers do to indicate that they are intertextually referencing other voices in their speech. However, the question of how the contextualization of playful intertextual media references specifically is similar or different from the signaling of reported speech in interaction is largely unanswered.

As for making playful talk recognizable, Glenn and Knapp (1987) find that 'unusual paralinguistic features such as marked changes in pitch, intonation, volume, or accent' can cue play. Both Tannen and Wallat (1987/1993) and Gordon (2002, 2008, 2009) show that high-pitched voice specifically indicates play and non-literal meanings. High-pitched voice is also used in teasing (Straehle 1993), a more specific type of play, along with cues such as intonation, stress, and vowel and voice quality, such as nasality. Overall then, scholars have identified pitch, intonation, and loudness as the most common cues used to make both intertextuality and play recognizable in interaction, with more specific features like accent and changes in voice qualities being used in particular instances. I extend the work on the animation of intertextual speech and play by examining the features used to make playful intertextual media references recognizable in interaction.

3. Methodology and data

I follow in the tradition of Interactional Sociolinguistics (IS), 'an approach to discourse analysis that has its origin in the search for replicable methods of qualitative analysis that account for our ability to interpret what participants intend to convey in everyday communicative practice' (Gumperz, 2015: 309). The data for this study is from a larger corpus of 45 h, 24 min of 40 digital audio recordings of everyday conversations over one year (August 2014–August 2015) primarily among white European American adult friends, myself included, of the Millennial generation in their mid to late twenties. Using a

small and unobtrusive digital recorder, or sometimes my mobile phone, ensured that I could easily record long stretches of conversations in various settings, such as in private homes and restaurants. All participants consented to being recorded and have been given pseudonyms here. I recorded conversations at fairly frequent intervals at my partner Dave's house and with other friends. In addition to Dave's housemates and their friends, I also recorded conversations with friends in the absence of Dave, which took place on a university campus. I ultimately chose five conversations of approximately 7 h of talk among 10 friends to analyze in this study, selecting these five primarily due to the relatively high concentration of media references they contained. Three of the conversations were recorded at Dave's house where he lived during that year with three friends: Todd, a roommate from college who worked as a software developer; Jeff, Todd's coworker; and Paula, a graduate student in linguistics in the same program as Dave and me (all participants except me are given pseudonyms). I did not live at the house but was present during the recordings. One conversation was recorded at a diner and another in a graduate student lounge.

As Tannen (1984/2005: 43–44) observes, 'recording a conversation among friends that would have taken place anyway makes available for study patterns of language use that do not emerge among strangers, such as playful routines, irony and allusion, reference to familiar jokes, and unstated assumptions.' These playful patterns of interaction are my primary analytical focus. These speakers playfully referenced books, films, TV shows, songs, YouTube videos and Internet memes. Our shared media knowledge is reflected in unstated assumptions which ultimately contribute to building and sustaining relationships and community based around knowing the same media references.

If I had questions about the references or my interpretations, I asked the participants for their insights through 'playback' interviews (e.g., Labov and Fanshel, 1977; Tannen, 1984/2005), supplementing these with additional emails, text messages, and Facebook chat messages throughout the analysis process. It was often the case that participants provided me with media sources that they had referenced but which I did not know, and in other instances they gave me insights into possible explanations for their own and each other's behavior. Tannen (1984/2005) asserts that 'playback is the litmus test of interpretation' (49), serving as an indicative check of the researcher's interpretations. Following Tannen's methodology then, I recorded, transcribed, studied, and made interpretations about the data, and then did playback and follow-up with participants, thereby checking, revising, and expanding my analysis.

For the analysis presented in this paper, I coded the signaling mechanisms (based on my own auditory perception, similar to what the participants would have experienced in the actual interactions) that speakers employed in quoting 148 media references across 5 conversations. Almost all of the media references in my data set (141/148; 95%) are null quotatives and are signaled instead with contextualization cues. The fact that most of these references do not involve quotatives (such as 'be like,' 'like' and 'say') underscores that they are understood to be shared texts, with no need for a quotative. Indeed, D'Arcy (2012) writes that the null form, (along with *be like* and *go*) favors quotes with "mimetic encoding", defined by Buchstaller (2008: 23) as "quotes that contain sounds or voice and gestural effects." My data and its analysis are in line with this finding, showing that zero quotatives make use of other audible contextualization cues.

I have not done close phonetic analysis of these cues, in large part because the recordings were made in noisy settings where dishes clanked, cats meowed, and one participant in particular (Todd) frequently whistled songs in the background. Furthermore, the contextualization cues sometimes occur in moments of overlapping turns of talk among different speakers, making a systematic phonetic analysis difficult, if not impossible. Additionally, since I worked from only audio-recorded data, it is possible that embodied contextualization cues also occurred that I was not able to account for here (for work on non-verbal embodied cues used to signal intertextuality, see Clark and Gerrig 1990; Streeck 2002; Sidnell 2006; Earis & Cormier 2013; Thompson and Suzuki 2014; Blackwell et al., 2015; Stec et al., 2015).

It is extremely unusual for a media reference in my data to not be accompanied by a marked contextualization cue at all. Only three references in 148 had no identifiable contextualization cue that seemed to mark the reference as unique from the unfolding discourse stream. In one of these cases, the initial use of the reference did not garner any listener response, but when the same speaker repeated the reference later with a signaling mechanism, listeners did recognize and respond to the reference (see Sierra 2021 on a close analysis of recipient responses to media references). This provides strong evidence that contextualization cues do extra work to highlight media references for listeners, thus making the cues used to signal media references worth further consideration.

4. Five ways media references are made recognizable in talk

4.1. Word stress and intonation

The most common contextualization cue used to signal media references in my data is word stress (123/148; 83%), consisting in different instances of vowel lengthening, loudness, and a pitch accent, or raised pitch (see Bolinger 1958; Zsiga 2012). Word stress highlights the media reference for interlocutors (see also Tetreault 2009) by making it stand out in the speech stream. A listener might hear this word stress within the media reference itself, or in a word preceding the media reference. Similarly, reproductions of particular intonation contours lifted from media texts are often distinguishable by virtue of the particular placement of word stress. These findings are reminiscent of the key prosodic features Selting (1994) identifies in creating emphatic speech styles. Here however, they occur primarily as a function of mimicking media texts and making them recognizable in talk (a finding also mentioned in a single instance by Tetreault, 2009), which in turn might create the perception of a more involved and emphatic speech style.

In the following example, four of us have been talking in the kitchen at Dave's house. Paula, Todd, Dave and I are discussing the apparent nonchalance of Todd's friend, Aaron, during a camping trip where he had vomited in a tent that we were all sharing. Here Jeff makes the same media reference twice. The first time, he does so without any contextualization cues and is overlapped with speech from Paula. This results in nobody responding to his reference. However, the second time, he is not overlapped and uses a particular word stress as a contextualization cue. This results in listener uptake of his reference (in this excerpt, like those that follow, media references most relevant to the analysis are indicated by a combination of quotation marks and underlining.)

(1) (Sierra, 2021:117–118)

1 Sylvia I love though how like the whole next day we were just-
2 making fun of hi:m,
3 and like, I- and he like didn't ca(h)[r(h)e(h)
4 Paula [He's v-
5 [I was thinking the same thing he's very stoic.
6 Todd [He- he does NOT-
7 He does^not care.
8 Paula He's just like-
9 [He was just-
10 Jeff [“Long hair don't care.”
11 Paula He [just like^o:wns it he's like,
12 Sylvia [He's just like “Yep, this happened.”
13 Paula “Yeah, psh” =
14 Dave = Yeah?
15 Paula “You know it.” =
16 Dave = Haha [haha.
17 Paula [“Yup.
18 Todd [(???)
19 Paula “That's [about what happened.”
20 Todd [(doesn't give a shit.)
21 Paula Yeah.
22 No that was very..very impressive.
23 Jeff “Lo:ng hair^do:n't care.”
24 Paula [“Lo(h)ng [h(h)air [do(h)n't ca(h)re.”
25 Sylvia [Hahaha.
26 Todd [Aaron^I:S “long hair don't care.”

In line 3, I laughingly observe how Aaron “didn't care” that we were making fun of him the next day, and Paula aligns with me in line 5 by agreeing that “he's very stoic”. Then Todd aligns with Paula and I in lines 6–7 by affirmatively stating “he does not care.” This talk, perhaps along with Paula's quotative “He's just like-” in line 9, is what presumably triggers Jeff to make a meme reference in line 10, saying “Long hair don't care.” This phrase became text for a variety of Internet memes following American R&B singer/songwriter Lloyd's #1 hit single in 2007 “You”, featuring Lil Wayne, singing the lyric, “Me & lil' Lloyd baby, long hair don't care”. After this, the phrase spread to long-hair forums, Instagram, Pinterest, and other websites in the form of images of people and characters with long hair accompanied by the phrase “Long hair don't care”.

Jeff, the only friend present who did not go on the camping trip, becomes involved in the talk by referencing the Internet meme “Long hair don't care” to comment on Aaron's hair style and behavior. When Jeff first references this meme in line 10, no one seems to hear him or pick up on it. Instead, there is overlapping constructed dialogue for Aaron in lines 12–19, furthering the description of his laid-back attitude in the face of embarrassment. But when the talk slows down a bit, Jeff persists in making the meme reference again in line 23. This time, Jeff uses vowel lengthening and a slightly higher pitch, on ‘don't’; this gives an exaggerated stress pattern to the phrase, highlighting the intertextual work Jeff is doing. Unlike Jeff's first attempt to make this reference, now Paula and I demonstrate appreciation of the reference through Paula's repetition (line 24) and my laughter (line 25). Todd also states emphatically that “Aaron IS ‘long hair don't care’” (4), implying that Aaron embodies this expression. Both the lexical content of the reference and Jeff's use of stress make the reference recognizable for the other speakers. This example shows that stress may highlight media references for listeners. Jeff uses this meme reference as a way to become involved in the talk, and his utterance in line 23 initiates a sequence closing sequence (Schegloff 2007) which garners collaboration from his interlocutors in closing this sequence of talk. Inserting media references like this into conversations constructs Jeff's identity as a humorous, clever, and fun person. It also allows him to participate in conversations where he lacks the same epistemic access, or knowledge (in this instance, knowledge about the camping trip), as his interlocutors (see also Sierra 2021). Referencing this meme also allows Jeff and the rest of the speakers present to briefly bond over their shared enjoyment of this meme.

There are many other examples in my data set of speakers mimicking word stress as it occurs in the intonation contour of the referenced media. Speakers replicate specific intonation contours to depict the original media text being appropriated, especially when the humorous intonation contour itself is a critical part of why the media text is worth referencing. To this point, Zsiga (2012) notes that intonation can bring a referent into focus, and in many of the examples, there is something

distinctive about the original intonation contour of the media text which makes it quotable. In other words, many media texts themselves make use of a specific intonation contour to bring a referent into focus, which speakers attempt to reproduce in their media references.

In the following example, four friends are talking in the kitchen/dining room. Paula and I had been discussing the difficulty in defining 'intertextuality'. At the time, I was taking a graduate seminar on the topic, and Paula was writing a paper involving intertextuality for another graduate class. This is the earliest recording in my data set, recorded before I knew I would be studying media references. Our talk eventually led to wordplay around the word 'intertextuality' itself. In this example, Jeff (who is not a graduate student) becomes involved in the epistemically isolating talk (or talk which he does not have knowledge about) by making a media reference (see also Sierra 2021). He makes the reference recognizable through replicating the intonation contour of the original media text.

(2) (Sierra, 2021: 34)

1 Paula Hey, I have nothing against intertextuals.
2 Ok?
3 Dave Ha=
4 Sylvia =Haha [hahaha
5 Jeff [“Not that [there's anything ˆWRO:NG with that.”
6 Paula [They're just ˆpeople.
7 They're ˆju:st people.

Paula makes a pun using the word 'intertextual' as a stand-in for something like 'intersexuals' when she says, "Hey, I have nothing against intertextuals. Ok?" (lines 1–2). Jeff responds to Paula's joke by saying "Not that there's anything ˆWRO:NG with that" (line 5). Vowel lengthening, loudness, and a pitch accent on 'wrong' are all present in this media reference, and they function to repeat the exact intonation contour of the words being referenced – a phrase from the popular American television show *Seinfeld*. In Season 4, episode 17 of the show, a running joke begins where a newspaper journalist, and then others, are mistakenly led to believe that best friends Jerry and George are a gay couple. Jerry responds to the mistake by telling the journalist, "We're not ˆGA:Y! Not that there's anything ˆWRO:NG with that". Since the emphasis is on the word "wrong" in the original phrase, the phrase takes on an intonation contour consisting of a high pitch accent on "wrong" (also stressed via loudness and vowel lengthening) followed by a low falling tone (see also Bolinger 1958). Jeff uses the same intonation contour when he quotes *Seinfeld* in this example. Mimicking the source text's distinctive intonation contour along with the repetition of the wording of the line makes the media reference recognizable.

The fact that Paula keeps speaking in lines 6 and 7, rather than delaying her response, initiating repair, or indicating in any other way that Jeff's utterance was problematic for her confirms that this media reference was effectively signaled and thus understood. My laughter at Jeff's reference here also communicates that both Jeff and I have seen this particular *Seinfeld* episode. Thus, the reference temporarily serves to bring us together as friends who are familiar with the same TV show that was popular when we grew up in the 1990s. Once again, this example also shows how Jeff uses a media reference to become involved in epistemically isolating talk. In sum, stress is the most common way that speakers phonetically signal media references in my data set. Stress highlights the media reference as significant for the listener. Stress is often present in the media text itself, co-occurring with a particular intonation contour, and then this intonation is reproduced when speakers reference that media. Speakers commonly use this contextualization cue to make media references recognizable for interlocutors.

4.2. Pitch register shifts

Speakers also commonly signal intertextual media references by shifting to a markedly higher or lower pitch register (90/148; 60%) (see also Couper-Kuhlen 1996 on 'prosody mimicry'). Following Couper-Kuhlen and Selting's (2018) definition, pitch register refers to the location of a pitch configuration in a speaker's overall voice range; it can range from mid (the norm) to extra high or extra low. Part of the reason that a shift in pitch register is used so often in my data is because speakers are frequently voicing others who speak with a noticeably higher or lower pitch register than the speaker's baseline.

In the following excerpt, Dave had been trying to bring up a particular scholarly journal article, prompting me to recall a very specific media text—audio recordings I remembered having to listen to for a graduate course that Dave, Paula, and I had taken on phonetics and phonology. In this recorded media, our professor, whom we refer to as Lisa, had recorded audio files of herself saying the phrases "Did Maddy win the medal?" and "Maddy won the medal" with different intonation contours, which we had to download and transcribe. Paula did not remember the recording even though she took the same class as Dave and me.

In this excerpt, I was trying to recall the exact phrase that Lisa had recorded herself saying for the activity. Initially, Dave provides the intertextual reference in his typical pitch register, and I do not show evidence of having heard him or recognized his utterance as the reference for which I was searching. But then, Dave switches to a very high pitch register to imitate the professor's higher pitched voice while providing the correct reference. This time, I respond accordingly.

(3) (Sierra, 2021: 37)

1 Sylvia No, Lisa did those recordings that were like-
2 What were they, like-
3 Paula Which one-
4 Dave "^Ma:ddy bought [the medal."
5 Sylvia [“What is Mar- ^Mary gonna-”
6 Dave No: =
7 Sylvia =What the fu(h)ck wa(h)s-
8 Dave It wasn't the- [She wasn't lookin at the se^mantics.
9 Paula [Wha(h)t?
10 Sylvia No::
11 Dave (high pitch) “^MA:DDY GOT THE MEDAL.”
12 Sylvia YEAH, ^THA:T.
13 When we were in phonol- [yeah-
14 Dave (high pitch) [“^MADDY HAS THE ^ME:DAL.”

At first, Dave provides the phrase I am searching for, with “Ma:ddy bought the medal” (line 4) but in his own typical pitch register. I do not register that he is supplying the reference, however, possibly due to a combination of Paula overlapping with Dave, asking “Which one-” (line 3), Dave still using his typical pitch register, or because I was still searching my own memory for the right phrase. It is apparent that I do not recognize that Dave is supplying the correct reference because I still attempt to recall the phrase myself, saying “What is Mar-^Mary gonna-” (line 5), which prompts Dave to give an exasperated “No:” (line 6) and I laughingly begins to ask “What the fu(h)ck wa(h)s-” still not recalling nor recognizing the phrase.

Finally, Dave says the phrase again, this time very loudly and in a noticeably higher pitch register: “MA:DDY GOT THE MEDAL” (line 11). Dave's timing (with no overlap), loudness, and much higher pitch register draw attention to the phrase and effectively make it recognizable as the reference I had been searching for, and I respond accordingly, “YEAH, THA:T! When we were in phonol-yeah-” (lines 12–13) with no further hesitation, repair, or indication of confusion. Dave's pitch register shift and loudness here are crucial in effectively signaling a media reference, albeit a very in-group one, to a listener. This example shows how a noticeable shift in pitch register, alongside loudness in this specific case, can be used to make an intertextual reference recognizable, allowing Dave and I to recall and reminisce about a shared coursework media experience as graduate students.

4.3. Smile voice and laughter

Although knowing every time someone was smiling in my audio-recorded data is impossible, smile voice is typically auditorily perceptible due to lip retraction and mouth widening that occurs with smiling (Tartter and Braun, 1994). People likely smile more often than they actually produce laughter, and thus laughter often (but not always) coincides with smile voice in my data. 39% of the media references in my data set involved speakers using smile voice as they made a reference, and 26% of the references were made while laughing.

While examining smile voice and laughter, I observed that the women in the conversations I recorded use smile voice and laughter to signal media references more often than the men. Indeed, Goffman (1976: 48) observes that ‘it appears that in cross-sexed encounters in American society, women smile more, and more expansively, than men’ and Tannen (1996: 216) asserts that ‘ … in our culture, smiling is a sex-class linked behavior; in other words, women tend to smile more than men.’ Tannen's explanation for the reason that smiling is a sex-class linked behavior is that women are expected to smile more than men. She writes, ‘women are seen as severe and lacking in humor if they rarely smile, whereas men who do not smile often are far less likely to meet with negative reactions’ (1996: 217). It is not clear if Goffman and Tannen would have hypothesized that the phenomena of women smiling more than men could be extended to laughter, but my data indicates that it does, since women both audibly smiled and laughed more than men did when producing media references (gender and laughter reach significance at $p < 0.01$, see Table 1; gender and smile voice reach significance at $p < 0.01$, see Table 2). It would be interesting to see if men smile and laugh together at higher rates when no women are present, but I did not have the occasion to record a conversation with only men present. Yet I want to underscore than men *do* smile and laugh together occasionally in my data set, just at a lower frequency.

Table 1
A chi-squared table showing the relation between gender and laughter when making a media reference. The result is significant at $p < 0.01$.

	Laughter in Reference	No Laughter in Reference	*Row Totals*
Women	25 (15.92) [5.18]	37 (46.08) [1.79]	62
Men	13 (22.08) [3.73]	73 (68.92) [1.29]	86
Column Totals	38	110	148 (Grand Total)

Table 2
A chi-squared table showing the relation between gender and smile voice when making media references. The result is significant at $p < 0.01$.

	Smile Voice in Reference	No Smile Voice	*Row Totals*
Women	32 (23.46) [3.11]	30 (38.54) [1.89]	62
Men	24 (32.54) [2.24]	62 (53.46) [1.36]	86
Column Totals	56	92	148 (Grand Total)

In the example below, Dave and I are about to leave the house to buy chicken for dinner. Paula and I again reference the recording that Dave and I had referenced 2 min earlier (in example 3), of our professor, Lisa, saying "Did Maddy win the medal?" and "Maddy won the medal" with different intonation contours. Here, the reference to the recording is simultaneously intertextual and intratextual, following Hamilton's (1996) distinction, since it refers to the recording itself, which was heard many months ago, and to talk within the current conversation that occurred just a couple of minutes ago. I perform the reference drawing on the exaggerated intonation contours that Lisa had used in her recording, but I now use them for our current talk about buying chicken. Paula and Jeff join in, and as the intertextual and intratextual play grows increasingly absurd and metadiscursive, layering reference upon reference, the 'shared hilarity' (Chafe 1994) of the situation builds. Eventually, the two men present, Jeff and Dave, start to smile and laugh as they reference additional material. There is an additional transcription symbol introduced here to indicate smile voice; a smiling face symbol (☺) (following the conventions of Ericsson 2018).

(4) (Sierra, 2021: 44)

1 Dave You ready?
2 Sylvia Yeah. Should I bring my wallet or are you gonna buy the→
3 chicken?
4 Dave ^I'll buy the chick [en.
5 Paula [Ha! Hahaha(h)
6 Sylvia "Are ^you: gonna buy the chicken?" …
7 [Hahaha
8 Jeff ["Are you going to ^buy the chicken."
9 Paula [☺ "ARE <*harsh voice*> ^YOU GONNA BUY THE→
10 CHICKEN? ☺"
11 ☺ "ARE YOU GOING TO BUY THE ^CHICKEN?" ☺
12 Sylvia Ha.
13 Jeff "ARE YOU GOING TO ^BUY THE CHICKEN?"
14 Paula ☺ "ARE ^NO:T YOU GOING TO BUY THE CHICKEN?" ☺ =

Dave uses falling intonation in responding "I'll buy the chicken" (line 4) to my question, which was produced with rising intonation "Should I bring my wallet or are you gonna buy the chicken?" (line 2). This reminds both Paula and I of the preceding playful "Maddy won the medal" sequence with alternating intonation contours that had occurred just a couple minutes earlier. This is evidenced by Paula's laughter (line 5) and by my altered intonation and otherwise unnecessary repetition of the question, "Are you gonna buy the chicken?" (line 6) (mimicking "Did Maddy win the medal?"). This is also followed by my laughter, signaling the initiation of play with this mimicking of the earlier intonation contour. Paula and Jeff both respond by also playfully repeating this phrase with various intonation contours (lines 8–14), with Paula laughing and smiling throughout her contributions. The intertextual play here allows Paula and Jeff to become involved in Dave's and my epistemically isolating talk (see also Sierra 2016, 2021) about our dinner plans.

Next, yet another media text is referenced, and more smile voice and laughter are used in playfully layering another intertextual reference.

(5) (Sierra, 2021: 44)
15 Dave = "I wanna die [dot jpeg."
16 Paula [Hahahaha
17 Jeff Hahaha
18 Paula (h)(h)
19 Sylvia <*high-pitch*> ☺ "I wanna di:e" ha ☺
20 Paula Hahaha
21 Dave Hahahaha
22 Jeff ☺ "DO YOU WANT TO ^DI:E?" ☺
23 Sylvia Hahaha.
24 Dave ["I ^WA::NT TO DIE."
25 Jeff [☺ "Do you ^WA:NT to die?" ☺
26 Sylvia Hahaha
27 Jeff ☺ "Do ^yo:u want to [d-" haha ☺
28 Paula [(h) [<laughing> (h)
29 Dave [Ha. <yells> "^I:::! wa(h)nt→
30 to(h) di(h)e"

Dave, mocking exasperation with his friends' intertextual play, quotes an entirely different media source, saying "I wanna die dot jpeg" (line 15), which is an internet meme. This meme originally derived from an image of a dolphin leaping out of the sea in JPG (a format for compressing image files); behind the dolphin is a starry night sky with a rainbow accompanied by the text 'I wanna die' in comic sans, a casual font based on comic book lettering. The humor of the meme derives from the incongruence of the rainbow and dolphin with the bleak message conveyed in comic sans. This meme was often posted in Internet message boards simply via the text 'iwannadie.jpg', as a humorous reaction, which explains why Dave says the title of

the file name aloud here. Paula and Jeff respond to this reference with laughter (lines 16–17), while I repeat the reference in a high pitch register and with smile voice (line 19), causing Paula and Dave to laugh (lines 20–21). Then Jeff begins to use smile voice with his references, repeating the intonation contour that we had been using for "Did Maddy win the medal" and asking loudly with smile voice, "DO YOU WANT TO DIE" (line 22). He repeats this question with stress placed on different parts of the question two more times, smiling all the while (lines 25 and 27). Paula and I laugh throughout this (lines 23–28). In line 29–30, Dave laughs, before yelling, "I:::! want to die", breaking up with laughter as he does so.

Here we see Jeff and Dave using smile voice and laughter as they reference media, after Paula and I also do so. In this mixed-gender setting, all of the participants end up smiling and laughing together, enjoying the shared hilarity of this moment of escalating intertextual play. This example shows that smiling and laughter do not occur solely as cues for signaling media references, but also express shared enjoyment and humor derived from making such references. As we can see in this example, the references are also made recognizable via stress and intonation contours, loudness, pitch, along with the semantic content of the reference in the case of the 'iwannadie.jpg' reference. While speakers smile and laugh when making media references according to gendered expectations, they also do so, regardless of gender, for rapport, group solidarity, alignment, and enjoyment of the conversation.

4.4. Performing stylized accents

The performance of marked stylized accents (differing from the speakers' typical pronunciation in their own dialects) occurs in 31% of the media references in my data (see also Sierra 2019, 2021). Rampton (2009) defines stylization as involving "reflexive communicative action in which speakers produce specially marked and often exaggerated representations of languages, dialects, and styles that lie outside their own habitual repertoire" (149). Speakers in my data set perform stylized accents when the source text they reference involves an accent noticeably different from the speaker's typical pronunciation. This only happens occasionally, likely because the speakers in my data generally consume media that is spoken in English and in most cases, in Standardized American English (SAE) (see also Sierra 2019). However, my data includes many notable examples where speakers use 'depictive delivery' (Clark and Gerrig 1990), utilizing specific vocal attributes to voice characters from media who speak in accents that are distinct from SAE.

The vast majority of the examples in my data featuring stylized accents are made by one speaker: Jeff. Jeff makes almost $^{3}/_{4}$ of the total number of media references performed with stylized accents. Half of the references he makes include stylized accent performances. Paula, in comparison, makes about $^{1}/_{4}$ of the total media references coded for stylization, and she performs a stylized accent in 20% of her media references. Todd and I each only perform stylized accents in two of our media references. These numbers point to individual conversational style differences (Tannen, 1984/2005). Jeff as an individual is, in the first place, a prolific media quoter, contributing around one third of all the media references in the data set, and he also performs half of his media references with stylized accent performances.

In the data set, Jeff performs British, French, Indian, Polish, Russian, and Spanish accents. Why does he perform so many stylized accents? Jeff explained to me in playback that as a child of an Egyptian and Lebanese father and an American mother of German, Irish, and Czech descent, he grew up in international circles in Washington, D.C. He believes his upbringing and exposure to many foreign languages and their English accents might have led to him developing a familiarity with various accents. He also told me that he realized at an early age that performing stylized accents made others laugh, and this became an incentive for this kind of linguistic performance. He also once took an improv acting class, where he practiced such performances. Here it is important to acknowledge that these playback responses are indicative of later reflection on the linguistic behavior that I recorded, and are simply one individual's interpretations of their own language use.

The excerpt I present below follows example 2, where Dave and I were reminiscing about the recording that our professor Lisa had us listen to and transcribe in order to learn about intonation contours in her phonetics/phonology course. Here, Jeff (again, not a graduate student) becomes involved in the epistemically talk by referencing the 2003 film *The Room*, voicing Polish actor Tommy Wiseau's character, with vowel lengthening, intonation mimicry, lowered pitch register, and a stylized performance of Wiseau's accent. This is actually the third time in the conversation Jeff has attempted to make this exact same reference, but I present the following excerpt because here his contribution is finally acknowledged by the group, likely due to the timing of the reference.

(6) (Sierra 2021: 49)

1	Dave	Maddy had mad maddy mad mad mumuhuhmuhmuh
2		[madmamadmadaaamamama
3	Paula	[muhmuhmuhmuh [hahaHA
4	Sylvia	[and we just had to listen to [Lisa over and over again.
5	Paula	[(h) "PORK CHOP ->
6		^SA:NDWICHES!"
7	Dave	Ha yeah.
8	Paula	That's hilarious.
9	Sylvia	[And it was like "no:!"

10 Jeff (*low pitch, with Wiseau accent*) [“You’re breaking my ^hea:rt Lisa:!”
11 Sylvia Ha [haa!
12 Dave [Haha [ha!
13 Paula [Ha [haha!
14 Jeff [“You’re tearing me apart, [Lisa.”
15 Sylvia (*low pitch, with Wiseau accent*) [“You’re tearing me aPA:RT!” ha

My mention of how “we just had to listen to Lisa over and over again” (line 4), referring to the professor who taught Paula, Dave, and I, provides an opportunity for Jeff to join the conversation by referencing a scene in *The Room* where Wiseau dramatically yells to his fiancée, also named Lisa, “You are tearing me apart, Lisa!” Jeff signals the reference by repeating its words along with its vowel lengthening, overall lowered pitch (with a slight pitch accent on ‘heart’), and a stylized accent performance mimicking Wiseau, saying “You’re breaking my ^hea:rt Lisa:!” in line 10 (Wiseau is thought to be from Poland). Whereas the previous two times Jeff made this reference, none of the group responded or acknowledged his contribution to the conversation, this time, everyone laughs (lines 11–13), demonstrating that the reference is effectively made recognizable here (although Paula admitted in playback that she had not seen *The Room* but was laughing at the general silliness of the utterance; Dave and I had seen the film). This reference allows Jeff to participate in the conversation, and it brings the group together in this moment of shared hilarity, which occurs since most of us (excluding Paula) had epistemic access to *The Room*, having seen it previously (see also Sierra 2021).

In sum, performing stylized accents can be a contextualization cue for making a media reference recognizable, particularly when the character or singer being repeated speaks with an accent markedly different from the speaker’s typical pronunciation. Certain individuals, like Jeff, perform stylized accents much more than other speakers. While the performance of stylized accents constructs certain others as ‘different’ and also differentiates participants, it also reinforces sameness among interlocutors in that it is jointly recognized as a strategy for producing a media reference in an accent not typically employed by the group, and the knowledge of the media text is often shared.

4.5. Singing

Similar to the performance of a stylized accent as a contextualization cue to signal media references, singing only serves to make media references recognizable in my data when the source text calls for it. Singing itself is somewhat marked in conversation (see Frick 2013; Stevanovic and Frick 2014, Azios & Archer 2018), and so when it does appear, it often references a song assumed to be known to other participants in the conversation. Singing occurs when speakers invoke the melody and/or words of a song, either from a film or TV show, or from a stand-alone song (e.g., a song produced by a musician or group for radio play, but not as part of a film or TV show). Only 22 media references in my data set involve singing.

In the next example, Paula and Todd’s kitten, Hydra, has just leapt to the top of the refrigerator in the kitchen. This had become rather expected and not unusual, although it was still relatively new at this time, which is demonstrated by how I comment on it below. There is an additional transcription symbol introduced here to indicate singing; a musical note symbol (♫) (following the conventions of Frick 2013 & Warnock, 2015).

(7) (Sierra 2021:56)

1 Sylvia (*sing-song intonation*) There she goes.
2 Paula “Tha:r sh(h)e [^blo:ws” haha
3 Sylvia [Hahaha
4 Jeff [♫“There she go:es again.”♫
5 Sylvia ♫“The:re [she: [go:es,”♫
6 Todd [I want one of those (???)
7 Paula ♫[“go(h):es,”♫
8 Sylvia ♫[“Cli:mbing o:n the fridge” haha♫
9 Paula ♫[“There she go(h)es”♫ hahahahahaha
10 ♫“On the [fridge agai:n”♫
11 Sylvia [Hahaha

I remark on Hydra’s jump onto the fridge, saying, “There she goes” (line 1) in a sing-song intonation, with a high pitch accent followed by a low falling tone over the word “goes”. Paula builds on my contribution after a slight pause when she quotes the 1851 book *Moby Dick* by Herman Melville, which tells the tale of a whaler captain’s quest for a white whale. Paula laughingly says, “Thar she blows” (line 2), which commonly refers to sighting a whale. Jeff is the first one to reference the song ‘There She Goes’ by The La’s, saying, “There she goe:s again” with a slight musical quality to it. Then I reproduce the song, mimicking the melody of it by singing it with vowel lengthening in each word, and increasing the pitch with each word: “The:re she: go:es” (line 5). Paula joins in laughingly with “go(h):es” (line 7), and continues with “There she go(h)es” (line 9), demonstrating both recognition of the reference and alignment by singing the same phrase with the same pattern of intonation (Stevanovic and Frick 2014, Azios & Archer 2018). As Paula continues to sing the original lyrics of the song, I modify the lyrics and map them over the original melody, singing instead, “Cli:mbing o:n the fridge” (line 8), causing Paula to laugh (line 9). Then Paula collaborates in the joint activity of creating new lyrics for the song with “On the fridge agai:n” (replacing the

original lyrics of "There she goes again"). The fact that both Paula and I have epistemic access to this song and use it in this instance to mock a pet functions to bond us together as friends who grew up hearing this song on the radio and in popular films from our youth that featured it in their soundtracks, like *The Parent Trap* (1998).

Like speakers performing stylized accents to mimic referenced material, speakers often reference songs by singing them, as seen in this example. As Warnock (2015) observes, singing can contribute to an individual's clever, humorous, and playful persona. In the conversations I recorded, Paula sings the most references, followed closely by Jeff, and then Dave. I only sing once out of the 15 media references that I made (in the conversation above which had more instances of singing than any other). Paula, Jeff, and Dave might be more confident in their singing abilities than other speakers. Additionally, since Paula and Jeff are the only friends with improv and stand-up comedy experience, it is feasible that they would incorporate singing in their interactions more than the other friends to construct humorous and playful individual identities. Overall, singing is a noticeable strategy in the data set for referencing (sung) media.

5. Conclusion

In this paper, I examined how media references are signaled through specific contextualization cues in everyday talk across five conversations among friends. Building on Gumperz's (1977, 1982) and others' work on contextualization cues and related phenomena, this analysis has shown that media intertextuality is not only made recognizable by lexical repetition itself, but is also marked in the speech stream through prosodic, paralinguistic, and phonetic cues that have meaningful intertextual connections to media texts. In the excerpts I analyzed in this paper, friends introduce shared media texts into new contexts of everyday talk, creating intertextuality, which might be considered successful when other friends are able to 'get' the references. In order for listeners to pick up on the references, speakers use contextualization cues which signal media references within the ongoing stream of talk. Listeners cannot interpret media references if they lack access to the shared prior text, but even then, listeners cannot necessarily interpret a media reference by its content alone if it is not skillfully signaled in the speech stream by the speaker. The wealth of media references this friend group produces in the five conversations in this study allowed me to uncover the specific contextualization cues that speakers often use to contextualize media references.

I showed how speakers rely primarily on stress, often reproducing specific intonation contours from media in order to mark a reference as something unique in talk. I also demonstrated how marked pitch register shifts are often used to signal media references, particularly when speakers are reproducing a media text that was originally uttered in a pitch register significantly higher or lower than the speaker's typical pitch register. In addition, I showed that audible smile voice (and presumably smiling), laughter, the performance of stylized accents, and singing were used, albeit with less frequency, to make media references recognizable. I demonstrated that smile voice and laughter are linked to gender identities in the data, with the women tending to use smile voice and laughter more frequently than the men to signal their playful appropriation of media references. While the performance of stylized accents and singing are less common in my data set, I showed how certain individuals like Jeff and Paula might be more inclined to perform stylized accents or sing songs to invoke media texts which themselves include language varieties besides SAE and musicality. Thus, this paper builds on previous work on contextualization cues by applying this concept to the yet uncharted territory of making media references recognizable in talk beyond simple lexical repetition. At the same time, it illuminates our understanding of how intertextual processes unfold in talk.

While the data set exceeds 140 examples that were taken from five different conversations, it is important to acknowledge that this data set is still only a small glimpse of human interaction, among particular friends from a certain generation, within a specific age range, and with particular media consumption practices. All of these factors are contextually relevant, and while they condition the prolific use of media references, they also constrain the generalizability of the findings to some extent. Still, it is probable that the more common contextualization cues for media references that I identified—stress, intonation contours, and pitch register shifts—are likely to signal intertextuality more generally and media references more specifically in other instances of everyday interaction due to their high frequency in the data. While smile voice, laughter, stylized accent performances, and singing occur less frequently as contextualization cues for media references, they nonetheless provide interesting insights into gendered behavior (in the case of smile voice and laughter) and individual behavior (in the case of stylized accents and singing) in everyday conversation. These less common contextualization cues also illuminate constraints and affordances regarding the insertion of media texts into everyday talk when they are conditioned by portrayals of characters with distinctive accents or by singing. In this study I have provided a starting point to systematic analysis of some of the most common contextualization cues used to signal intertextual media references in everyday conversational discourse, which contributes to our understanding of intertextuality as an interactional phenomenon.

Appendix. Transcription Conventions

Punctuation reflects intonation, not grammar.
? indicates rising intonation at the end of a unit
. indicates falling intonation.
, indicates continuing intonation.
.. two dots indicate a noticeable pause
… three dots indicate a significant pause

=Equal sign indicates latching (second voice begins without perceptible pause)
[Brackets indicate overlap (two voices heard at the same time)
(??) indicates inaudible utterance
(h) indicates laughter during a word
"word" indicates a media reference
(words) indicates uncertain transcription
(*sound*) gives details about speech or non-speech sounds
ˆ indicates emphatic stress
CAPS indicates speech spoken loudly
: colon following a vowel indicates elongated vowel sound
- indicates an abrupt stop in speech; a truncated word or syllable
☺line continues
☺ word ☺ said with smile voice
♫ indicates singing

References

Auer, P., Di Luzio, A., 1992. The Contextualization of Language. John Benjamins Publishing, Amsterdam.
Ayaß, R., Gerhardt, C., 2012. The Appropriation of Media in Everyday Life. John Benjamins Publishing, Amsterdam.
Azios, J., Archer, B., 2018. Singing behaviour in a client with traumatic brain injury: a conversation analysis investigation. Aphasiology 32 (8), 944–966.
Bakhtin, M., 1981. The Dialogic Imagination. University of Texas Press, Austin TX.
Basso, E., 1992. Contextualization in Kalapalo narratives. In: Duranti, A., Goodwin, C. (Eds.), Rethinking Context: Language as an Interactive Phenomenon. Cambridge University Press, Cambridge, pp. 253–270.
Bauman, R., 1992. Contextualization, tradition, and the dialogue of genres: Icelandic legends of the kraftaskáld. In: Duranti, A., Goodwin, C. (Eds.), Rethinking Context: Language as an Interactive Phenomenon. Cambridge University Press, Cambridge, pp. 125–146.
Blackwell, N.L., Perlman, M., Fox Tree, J.E., 2015. Quotation as a multimodal construction. J. Pragmat. 81, 1–7.
Blyth, C., Recktenwald, S., Wang, J., 1990. I'm like, "Say what?!": a new quotative in American oral narrative. Am. Speech 65 (3), 215–227.
Bolinger, D.L., 1958. Intonation and grammar. Lang. Learn. 8 (1–2), 31–37.
Buchstaller, I., 2008. The localization of global linguistic variants. English World-Wide 29 (1), 15–44.
Buchstaller, I., D'Arcy, A., 2009. Localized globalization: a multi-local, multivariate investigation of quotative be like. J. SocioLinguistics 13 (3), 291–331.
Chafe, W., 1994. Discourse, Consciousness, and Time: the Flow and Displacement of Conscious Experience in Speaking and Writing. University of Chicago Press, Chicago.
Clark, H., Gerrig, R., 1990. References as demonstrations. Language 66 (4), 764–805.
Couper-Kuhlen, E., 1992. Contextualizing discourse: the prosody of interactive repair. In: Auer, P., Di Luzio, A. (Eds.), The Contextualization of Language. John Benjamins Publishing, Amsterdam, pp. 337–364.
Couper-Kuhlen, E., 1996. The prosody of repetition: on quoting and mimicry. In: Couper-Kuhlen, E., Selting, M. (Eds.), Prosody in Conversation: Interactional Studies. Cambridge University Press, Cambridge, pp. 366–405.
Couper-Kuhlen, E., 1999. Coherent voicing: on prosody in conversational reported speech. In: Ventola, E., Lenk, U., Bublitz, W. (Eds.), Coherence in Spoken and Written Discourse. John Benjamins Pub, Amsterdam/Philadelphia, pp. 11–34.
Couper-Kuhlen, E., Selting, M., 2018. Interactional Linguistics: Studying Language in Social Interaction. Cambridge University Press, Cambridge.
D'Arcy, A., 2012. The diachrony of quotation: evidence from New Zealand English. Lang. Var. Change 24 (3), 343–369.
D'Arcy, A., 2017. Discourse-pragmatic Variation in Context: Eight Hundred Years of LIKE. John Benjamins Publishing Company, Amsterdam.
De Decker, P., 2013. Phonetic shifting across narrative and quoted speech styles. Proceedings of Meetings on Acoustics 20 (1).
Duff, P., 2002. The discursive co-construction of knowledge, identity, and difference: an ethnography of communication in the high school mainstream. Appl. Linguist. 23 (3), 289–322.
Duranti, A., 1992. Language in context and language as context: the Samoan respect vocabulary. In: Duranti, A., Goodwin, C. (Eds.), Rethinking Context: Language as an Interactive Phenomenon. Cambridge University Press, Cambridge, pp. 77–100.
Duranti, A., Goodwin, C., 1992. Rethinking Context: Language as an Interactive Phenomenon. Cambridge University Press, Cambridge.
Earis, H., Cormier, K., 2013. Point of view in British Sign Language and spoken English narrative discourse: the example of "The Tortoise and the Hare. Lang. Cognit. 5 (1), 313–343.
Ericsson, S., 2018. The language of cisnormativity: children and parents in interaction with a multimodal app. Gend. Lang. 12 (2), 139–167.
Fletcher, J., 2005. Exploring the phonetics of spoken narratives in Australian indigenous languages. In: Hardcastle, W.J., Beck, J.M. (Eds.), A Figure of Speech: A Festschrift for John Laver. Lawrence Erlbaum, Mahwah, NJ, pp. 201–226.
Frick, M., 2013. Singing and codeswitching in sequence closings. Pragmat. Quart. Publ. Int. Pragmat. Assoc. (IPrA) 23 (2), 243–273.
Gaik, F., 1992. Radio talk-show therapy and the pragmatics of possible worlds. In: Duranti, A., Goodwin, C. (Eds.), Rethinking Context: Language as an Interactive Phenomenon. Cambridge University Press, Cambridge, pp. 271–290.
Glenn, P., Knapp, M., 1987. The interactive framing of play in adult conversations. Commun. Q. 35 (1), 48–66.
Goffman, E., 1976. Gender Advertisements. Harper & Row, New York.
Goodwin, C., Goodwin, M., 1992. Context, activity and participation. In: Auer, P., Di Luzio, A. (Eds.), The Contextualization of Language. John Benjamins Publishing, Amsterdam, pp. 77–100.
Gordon, C., 2002. I'm Mommy and you're Natalie": role-reversal and embedded frames in mother-child discourse. Lang. Soc. 31 (5), 679–720.
Gordon, C., 2008. A(p)parent play: blending frames and reframing in family talk. Lang. Soc. 37 (3), 319–349.
Gordon, C., 2009. Making Meanings, Creating Family: Intertextuality and Framing in Family Interaction. Oxford University Press, Oxford.
Gumperz, J., 1977. The sociolinguistic significance of conversational code-switching. RELC J. 8 (2), 1–34.
Gumperz, J., 1992. Contextualization revisited. In: Auer, P., Di Luzio, A. (Eds.), The Contextualization of Language. John Benjamins Publishing, Amsterdam, pp. 29–54.
Gumperz, J., 1982. Discourse Strategies. Cambridge University Press, Cambridge.
Gumperz, J., 2015. Interactional Sociolinguistics: A Personal Perspective. In: Tannen, D., Hamilton, H., Schiffrin, D. (Eds.), The Handbook of Discourse Analysis, 2nd editionWiley Blackwell, Massachusetts.
Günthner, S., 1992. Comments on Jenny Cook-Gumperz: gendered contexts. In: Auer, P., Di Luzio, A. (Eds.), The Contextualization of Language. John Benjamins Publishing, Amsterdam, pp. 223–232.
Günthner, S., 1999. Polyphony and the 'layering of voices' in reported dialogues: an analysis of the use of prosodic devices in everyday reported speech. J. Pragmat. 31 (5), 685–708.
Hamilton, H., 1996. Intratextuality, intertextuality, and the construction of identity as patient in Alzheimer's disease. Text Interdiscipl. J. Study Discour. 16 (1), 61–90.

Hanks, W., 1992. The indexical Ground of deictic reference. In: Duranti, A., Goodwin, C. (Eds.), Rethinking Context: Language as an Interactive Phenomenon. Cambridge University Press, Cambridge, pp. 44–76.
Hirschberg, J., Grosz, B., 1992. Intonational features of local and global discourse structure. In: Proceedings of the Workshop on Speech and Natural Language. Association for Computational Linguistics, pp. 441–446.
Jansen, W., Gregory, M., Brenier, J., 2001. Prosodic correlates of directly reported speech: evidence from conversational speech. In: ISCA Tutorial and Research Workshop (ITRW) on Prosody in Speech Recognition and Understanding.
Jones, T., 2016. AAE Talmbout: an overlooked verb of reference. Univ. Penn. Work. Pap. Ling. 22 (2), 11.
Klewitz, G., Couper-Kuhlen, E., 1999. Quote - unquote? The role of prosody in the contextualization Of reported speech sequences. Pragmatics 9 (4), 459–485.
Kotthoff, H., 2021. Parody and indexing gender: performing female comic figures on German TV. In: Linares Bernabéu, E. (Ed.), Gender and Identity in Humorous Discourse. Peter Lang, Berlin, pp. 13–28.
Labov, W., Fanshel, D., 1977. Therapeutic Discourse: Psychotherapy as Conversation. Academic Press, New York.
Mathis, T., Yule, G., 1994. Zero quotatives. Discourse Process 18 (1), 63–76.
Park, Y., 2009. Interaction between grammar and multimodal resources: quoting different characters in Korean multiparty conversation. Discourse Stud. 11 (1), 79–104.
Philips, S., 1992. The routinization of repair in courtroom discourse. In: Duranti, A., Goodwin, C. (Eds.), Rethinking Context: Language as an Interactive Phenomenon. Cambridge University Press, Cambridge, pp. 311–322.
Rampton, B., 2009. Interaction ritual and not just artful performance in crossing and stylization. Lang. Soc. 38 (2), 149–176.
Romaine, S., Lange, D., 1991. The use of like as a marker of reported speech and thought: a case of grammaticalization in process. Am. Speech 66 (3), 227–279.
Schegloff, E.A., 2007. Sequence Organization in Interaction: A Primer in Conversation Analysis, vol. I. Cambridge University Press, Cambridge.
Schwitalla, J., 1992. Comments on margret selting: intonation as a contextualization device. In: Auer, P., Di Luzio, A. (Eds.), The Contextualization of Language. John Benjamins Publishing, Amsterdam, pp. 259–272.
Selting, M., 1992. Intonation as a contextualization device: case studies on the role of prosody, especially intonation, in contextualizing story telling in conversation. In: Auer, P., Di Luzio, A. (Eds.), The Contextualization of Language. John Benjamins Publishing, Amsterdam, pp. 233–258.
Selting, M., 1994. Emphatic speech style - with special focus on the prosodic signaling of heightened emotive involvement in conversation. J. Pragmat. 22, 375–408.
Sidnell, J., 2006. Coordinating gesture, talk, and gaze in reenactments. Res. Lang. Soc. Interact. 39 (4), 377–409.
Sierra, S., 2016a. Intertextual Media References as Resources for Managing Frames, Epistemics, and Identity in Conversation Among Friends. Georgetown University. Ph.D. Dissertation.
Sierra, S., 2016b. Playing out loud: videogame references as resources in friend interaction for managing frames, epistemics, and group identity. Lang. Soc. 45 (2), 217–245.
Sierra, S., 2019. Linguistic and ethnic media stereotypes in everyday talk: humor and identity construction among friends. J. Pragmat. 152, 186–199.
Sierra, S., 2021. Millennials Talking Media: Creating Intertextual Identities in Everyday Conversation. Oxford University Press.
Spitulnik, D., 1996. The social circulation of media discourse and the mediation of communities. J. Ling. Anthropol. 6 (2), 161–187, 1996.
Stec, K., Huiskes, M., Redeker, G., 2015. Multimodal analysis of quotation in oral narratives. Open Ling. 1 (1), 531–554.
Stevanovic, M., Frick, M., 2014. Singing in interaction. Soc. Semiotic. 24 (4), 495–513.
Straehle, C., 1993. Samuel?" "Yes, dear?": teasing and conversational rapport. In: Tannen, D. (Ed.), Framing in Discourse. Oxford University Press, Oxford, pp. 210–230.
Streeck, J., 2002. Grammars, words, and embodied meanings: on the uses and evolution of so and like. J. Commun. 52 (3), 581–596.
Streeck, J., Hartge, U., 1992. Previews: gestures at the transition place. In: Auer, P., Di Luzio, A. (Eds.), The Contextualization of Language. John Benjamins Publishing, Amsterdam, pp. 135–158.
Thompson, S.A., Suzuki, R., 2014. Reenactments in conversation: gaze and recipiency. Discourse Stud. 16 (6), 816–846.
Tannen, D., 1986. Introducing constructed dialogue in Greek and American conversational and literary narrative. In: Coulmas, F. (Ed.), Direct and Indirect Speech. Mouton, Berlin, pp. 311–360.
Tannen, D., 1996. Researching gender-related patterns in classroom discourse. Tesol Q. 30 (2), 341–344.
Tannen, D., 1989/2007. Talking Voices: Repetition, Dialogue, and Imagery in Conversational Discourse. Cambridge University Press, Cambridge.
Tannen, D., 1984/2005. Conversational Style: Analyzing Talk Among Friends. Oxford University Press, Oxford.
Tannen, D., Wallat, C., 1987/1993. Interactive frames and knowledge schemas in interaction: examples from a medical examination/interview. Soc. Psychol. Q. 50 (2), 205–216.
Tartter, V., Braun, B., 1994. Hearing smiles and frowns in normal and whisper registers. J. Acoust. Soc. Am. 96 (4), 2101–2107.
Tetreault, C., 2009. Cité teens entextualizing French TV host register: crossing, voicing, and participation frameworks. Lang. Soc. 38 (2), 201–231.
Tovares, A., 2006. Public medium, private talk: Gossip about a TV show as 'quotidian hermeneutics'. Text Talk Interdiscipl. J. Lang. Discour. Commun. Stud. 26 (4–5), 463–491.
Tovares, A., 2007. Family members interacting while watching TV. In: Tannen, D., Kendall, S., Gordon, C. (Eds.), Family Talk: Discourse and Identity in Four American Families. Oxford University Press, Oxford, pp. 283–309.
Uhmann, S., 1996. On rhythm in everyday German conversations: beat clashes in assessment utterances. In: Couper-Kuhlen, E., Selting, M. (Eds.), Prosody in Conversation: Interactional Studies. Cambridge University Press, Cambridge, pp. 303–365.
Warnock, R., 2015. Dementia and Singing: A Conversation Analysis Case Study of Signing in Everyday Interaction. University of Colorado, Boulder. Ph.D. Dissertation.
Yule, G., Mathis, T., 1992. The role of staging and constructed dialogue in establishing speaker's topic. Linguistics 30, 199–215.
Zsiga, E., 2012. The Sounds of Language: an Introduction to Phonetics and Phonology. John Wiley & Sons, Hoboken.