4 Recipient design I: How do narrators mark quotation?

4.1 Introduction

Speakers in conversation very frequently present discourse that occurred in anterior situations. This phenomenon has been referred to under various headings: as a change in footing (Goffman 1981), speech reporting (e.g., Carter & McCarthy 2006), speech, thought, and writing presentation (e.g., McIntyre *et al.* 2004) and discourse presentation (e.g., McIntyre & Walker 2011), the term used in this study. 'Discourse,' as understood in 'discourse presentation,' casts the net wide to capture a broad variety of discourses including internal thought and emotion, non-verbal gesture, and external speech. For this latter category (external speech) further distinctions can be made: external speech can be habitual speech (the words someone uses or used to utter on similar occasions), hypothetical speech (the words someone might or might have uttered in some situation), and actual speech (the words someone purportedly uttered in a specific situation). This variability alone makes discourse presentation a multifaceted phenomenon. Yet two further aspects greatly add to the variability of the discourse presentation system: quotatives and reporting modes. Quotatives are typically verbs used to introduce forthcoming discourse presentation. As shown in Chapter 2, the quotative forms used in the NC comprise a small set of forms including SAY, GO, THINK and a few other forms. Reporting modes, on the other hand, are different ways of presenting discourse; in Chapter 2, I sketched the ten reporting modes annotated in the NC. It was shown that direct quotation, also referred to as constructed dialog (Tannen 1986, 1988), in its two incarnations as Free Direct and Direct, is by far the most frequent report type in the NC. Given its dominance in quantitative terms and its centrality in turning storytelling into drama (Tannen 1986, 1988) this mode will be of particular interest in, not only this present, but also the following chapter. The chapters will view the use of discourse presentation as a means by which narrators, following the principle of recipient design, attend to their interlocutors' processibility needs (Chapter 4) and, respectively, intensify their interlocutors' interest in the telling (Chapter 5).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

Specifically, my intentions in the present chapter are three-fold. In Sections 4.2 and 4.3, the aim is to examine how narrators *mark* discourse presentation; the focus will be on the marking of Free Direct and Direct discourse presentation (referred to as quotation or dialog), since it is here that the change in footing causes what I will call the 'boundary problem,' that is, the difficulty for the hearer to distinguish discourses attributed to different speakers and the concomitant necessity for the narrator to demarcate the beginning and also end of quotation as a sort of processing instruction for the recipient. While Section 4.2 is concerned with interjections occurring in first position (that is, as the very first word) in the quote and their potential function as quotation markers, in Section 4.3, the focus is on short silent pauses and their potential function as a 'quote-unquote' signal.

In the remainder of this introduction to the present chapter though I wish to make a few notes intended to deepen the understanding of the discoursal and interactional underpinnings involved in doing discourse presentation in narrative. Specifically, I will introduce three notions that I believe are key in the context of discourse presentation: immediacy, demonstration, and the above-mentioned boundary issue.

Immediacy

The use of discourse presentation effects a discoursal intertwining of two distinct discourse situations (see Short *et al.* 1996: 112): a (mostly) anterior, told, situation (comprising its spatiotemporal setting, the speakers and their communications) gets approximated toward a present, telling, situation (comprising the present spatiotemporal setting, the present speakers along with their present communications) in such a way that the two situations come *more or less* to overlap discoursally.

The approximation is a gradient phenomenon: the degree to which the two discourse situations can be overlaid varies depending on which discourse presentation type is chosen. I will argue in the following that the degree of overlay is in direct correlation with the degree of *immediacy* of the presentation type chosen.

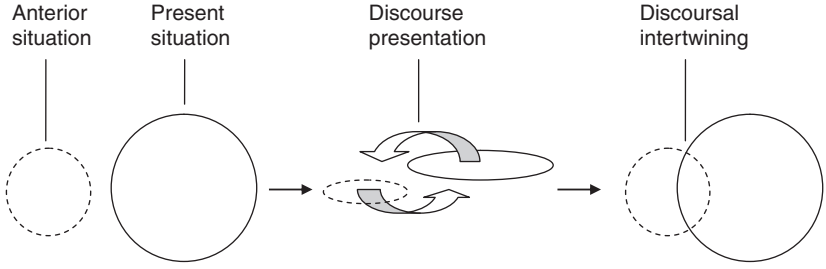


Figure 4.1 Discoursal intertwining as an effect of discourse presentation

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

Leech & Short (1981) assumed that discourse presentation types could be aligned along a cline of narratorial interference, viz. in terms of the degree to which the narrator apparently interferes with, or is in control over, the discourse presented (Leech & Short 1981: 324).[1] Their model of narratorial interference has been highly influential in research on discourse presentation, (e.g., Short *et al.* 1996, Semino & Short 2004, McIntyre *et al.* 2004, Short 2007, McIntyre & Walker 2011). Given that Leech & Short primarily examined discourse presentation in fiction it is not surprising that the model takes the perspective of the narrator. Alternatively, discourse presentation, particularly in conversation, where interaction between participants looms large, can be examined with a focus on what is achieved by it for the recipient. Seen from the recipient's viewpoint, another cline can be discovered: that of immediacy.

Immediacy derives from Goffman's (1981) analytical framework of 'production format' and the choices among *internal* speaker roles it affords. Goffman distinguishes, inter alia, 'author' and 'animator,' where author is understood as "someone who has selected the sentiments that are being expressed and the words in which they are encoded" (Goffman 1981: 144), while animator is conceptualized as someone "mov[ing] his lips up and down to the accompaniment of his facial (and often bodily) gesticulations, and words can be heard issuing from the locus of his mouth" (Goffman 1981: 145).[2] Clearly, under many (perhaps most) circumstances the two roles are one: "the individual who animates is formulating his own text" (Goffman 1981: 145) and the speaker acts in the double role of author-animator. In discourse presentation, however, the two roles do not simply overlay. While a presenter, as any speaker, is, by definition, always in the role of animator, they need not always and/or not fully perform the role of author, understood in Goffman`s sense, of the discourse animated. The type of discourse presentation where the two roles seemingly drift apart most is direct quotation, in its incarnations as Direct and Free Direct: there, the presenter seems to animate, not their own words, but the words of some story character. The author-animator unity collapses: the presenter acts as 'animator-only' but not as author-animator.

While this account is accepted widely (e.g., Holt & Clift 2007, Brendel *et al.* 2011), it requires a conceptual extension. For the choice between animator-only and animator-author by definition builds on 'pretense', a notion discussed at large by Searle (1975) in the context of the illocutionary acts performed in fictional writing. Anticipating the reservations (that there may not be an original utterance and that quotations are selectively depictive rather than faithfully descriptive), presenters, acting in the role of animator-only, inevitably 'co-author' the quoted wording such that the exact locutions in the anterior situation were different ones.[3] In formulating the quote, the presenter *is*, in actual fact, both animator and (co-)author of the quoted utterances. But the utterance is presented *as if* no authorial interference

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

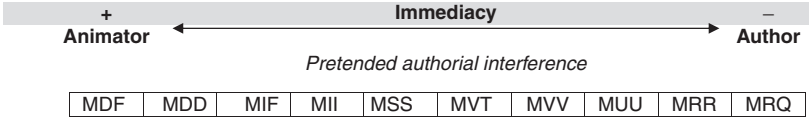


Figure 4.2 Cline of immediacy as a function of pretended authorial interference (see Leech & Short 1981); MDD: Direct, MDF: Free Direct, MII: Indirect, MIF: Free Indirect, MSS: Representation of Speech Act, MVT: Voice with Topic, MVV: Representation of Voice, MUU: Representation of Use, MRR: Anaphoric Reference to Discourse Presentation, MRQ: Request for Discourse Presentation

with its wording had occurred. The *effect* the pretended dissociation from the role of author has on the recipient is decisive: if the presenter seemingly 'steps behind' the characters merely lending their voice to them but letting them speak for themselves, the *illusion* is created that "the incident is presently occurring" (Mayes 1990: 346). Given this illusion, two additional, virtual, participant roles emerge: the reported speaker becomes a virtual addressor and the recipient a virtual addressee (see Section 1.3). As virtual addressees, recipients 'witness' the reported discourse (see Holt 1996: 236), and, as witnesses, they have *immediate* access to the reported utterance. In terms of the above mentioned discoursal intertwining, the overlay of the reporting situation and the reported situation is complete.

Crucially, in presenting discourse, the choice of animator-only and animator-author is not an either-or choice: the pretended dissociation from the role of author can be more or less. As a result, the recipient's access to the discourse reported can be more or less immediate. In Figure 4.2, I present a model of discourse presentation where discourse presentation types are ordered on a cline of immediacy understood as a function of pretended authorial interference.

To remind the reader of the differences between the categories, the following examples may be helpful illustration; only the example for MDD is authentic, all others are invented.

MDF	Free Direct	**Can I have my videos?**
MDD	Direct	**Wayne said, can I have my videos?** (KE5-N1)
MIF	Free Indirect	**Could he have his videos?**
MII	Indirect	**Wayne said, can he have his videos?**
MSS	Representation of Speech Act	**Wayne asked for his videos.**
MVT	Representation of Voice with Topic	**He talked about his videos.**
MVV	Representation of Voice	**He sat there talking.**

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

MUU	Representation of Use	**He said "vids".**
MRR	Reference to Discourse Presentation	**He really said that.**
MRQ	Request for Discourse Presentation	**What did he say?**

Free Direct (MDF), on the extreme left end of the cline, represents the category in which the reporting speaker's pretense *not* to formulate his/her own text but instead to merely animate that of a character is maximal. As Leech & Short put it, in Free Direct, where the quote is not announced by a reporting clause, a situation arises as if the narrator "has vacated the stage and left it to his characters" (Leech & Short 1981: 334). The presented utterance is maximally immediate to the recipient.[4] In Direct (MDD), the category next to MDF on the cline, the presence of the reporting clause begins to assert the speaker's rights as author: while the report itself is fully ascribed to the non-present speaker whose discourse is animated, the introduction of the report by the reporting clause is a reminder to the recipient of the presenter's potential to interfere with the report. Given this reminder (and its close structural association with the quote[5]), the quoted utterance is slightly less immediate to the recipient. In Free Indirect (MIF), the third category on the cline, the utterance presented undergoes a decisive authorial revision: all deictics anchoring the quoted utterance in the anterior speech situation are deleted and replaced by deictics conforming to the presenter's deictic system. Thereby the presented discourse is transposed from the spatiotemporal coordinates of the anterior context of utterance to those of the present context of utterance. That is, in MIF, the animator-author balance is shifted even more towards the presenter as author; consequently, the recipient's access to the original utterance is further restricted and immediacy is further decreased. Indirect (MII), the subsequent category on the cline, strengthens the authorial role of the presenter in that MII is distinguished from MIF by the use of a reporting clause to announce that discourse presentation is forthcoming. As with Direct (MDD), this announcement is an addtional authorial intervention and has the effect of further constraining the recipient's access to the original utterance.

The subsequent categories on the cline on the right of MII share an important structural property. They are distinguished from the Direct and Indirect categories by the fact that instead of a bi-clausal structure consisting of reporting clause (such as *he said* and *she goes*) and reported clause (the report itself), the presentation is expressed in one clause. The effect of this structural reduction is that the discourse presented is no longer somewhat detachable from the presenter's own discourse: it has completely merged into it. The mono-clausal structure greatly increases the presenter's 'authorship' and, accordingly, diminishes the recipient's access to the discourse presented. The first such mono-clausal category is Representation of Speech

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

Act (MSS). In MSS, the presenter presents his/her *interpretation* of the illocutionary force of the utterance. With the report effectively encoding the outcome of an analytical process, the presenter now asserts full and sole authorship rights. On the other hand, it is essential to note that MSS shares with the direct and indirect categories the property that it facilitates recoverability of the utterance reported. While the direct categories present the utterance using a depictive locution (from which the recipient can derive the illocution), the indirect categories provide sufficient information primarily on the content but also the form of the utterance so that possible locutions and/or illocution can be constructed by the recipient; MSS, on the other hand, directly presents an analysis of the illocution (from which possible locutions can be derived).[6] Given that, thus, MSS supports the retrievability of the underlying utterance (in whatever possible locution), the immediacy to the recipient afforded by MSS is still greater than the immediacy supported by the remaining categories, none of which support retrievability. There, the reported utterance cannot be re-constructed by the recipient either in terms of locution or illocution. As a result, the discourse's immediacy to the recipient sees further decisive reductions.

The categories include Representation of Voice with Topic (MVT), Representation of Voice (MVV), Representation of Use (MUU), Anaphoric Reference to Discourse (MRR), and Request for Discourse Presentation (MRQ). As noted in Section 2.2.2, both MVT and MVV refer minimally to speech without specification of illocution or locution. The two are distinguished in terms of content description: while MVT states the discourse's topic, MVV only states that discourse occurred (without any indication of its aboutness), which causes the discourse to further lose in immediacy to the recipient. The next category, Representation of Use (MUU), is decidedly selective, picking out from the utterance a single linguistic or paralinguistic peculiarity (a vocable, intonation, voice quality, dialect, etc.) and its focus is meta-linguistic, serving to expose the feature *as* a peculiarity. The recipient's access to the discourse *as a whole* is drastically impoverished. The final two categories, Reference to Discourse (MRR) and Request for Discourse Presentation (MRQ), which are boundary types bordering on narration, have in common that the utterance reported is only being referred to. The discourse is thus minimally immediate to the recipient and, as regards discoursal intertwining, the reported situation is maximally distant from the reporting situation.

In sum, reporting types can be ordered on a cline of immediacy, with the reported discourse's immediacy depending on variation in production format (see Goffman 1981): the more the reporter is purportedly in the role of animator, lending their voice to a non-present speaker, the more the reported discourse is made immediate to the listeners – immediate in the sense of creating the illusion as if the non-present speaker whose discourse is being presented were immediately present producing the discourse reported

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

themselves. Presentation types high in immediacy breathe life into the characters on that stage: they (seem to) speak for themselves through the presenter. In presentation types low in immediacy the presenter acts as a sort of stage manager talking *about* the characters and their discourses, pointing to them on the stage and describing their discourses, with varying degrees of explicitness and completeness, but not animating them. The discourse remains at a distance.

The implications of immediacy for this study, whose overarching theme is the co-construction of narrative, may not be immediately obvious. As noted in the Introduction, co-construction is not only achieved by recipients' contributions to storytelling but also by narrators. Their co-construction is implicated in recipient design (see Section 1.7), that is, in the strategy to construct narrative *on the basis of* the narrator's recipient model including a model of their needs, interests, and resources and *with an orientation* to these needs, interests, and resources (see, for example, Sacks 1992: 540). Now, immediacy, in the above-described sense, is directly oriented to the recipient's *interests*. This is particularly noticeable where maximal immediacy is facilitated, viz. in the use of constructed dialog in its two incarnations as Free Direct and Direct. As noted, maximal immediacy creates the illusion of the reported speaker being present in the telling situation producing the reported locution themselves. Seemingly present, the reported speaker becomes a sort of participant in the telling situation, what I referred to as a virtual participant (see Section 1.3). In that it is uttered by a virtual participant, the reported utterance is virtually addressed to the physical participants co-present in the telling situation, including first and foremost the recipients, who, on being thus addressed, become virtual addressees. Thus, the use of constructed dialog entails virtual participation, including the virtual addressees (the story recipients) and the virtual addressor (the reported speaker). Though only virtual, the relationship that obtains between them is not dissimilar to the relationship between interlocutors in normal turn-by-turn talk, where turn allocation techniques are used including 'current speaker selects next speaker' and 'next speaker self-selects' (see Sacks *et al.* 1974) and where, hence, "the obligation to listen" (Sacks 1992: 41) becomes crucial in order to be able to respond coherently and with no gap/no overlap (see Sacks 1992: 649–650). The reported utterance is for the recipients virtually as relevant *as if* they had to respond to it, thus entering into a turntaking relationship with the virtual addressor. True, this does not very frequently happen: story recipients rarely 'respond' to the virtual addressor's utterance but it does happen (see Section 6.3, which reports on a case study on dialog constructed by the recipient, most of which is indeed in response to the virtual addressor's utterance). The point, however, is not whether or not it happens. The point is that even where it does not happen, as in most cases, the virtual turntaking relationship between addressor and addressees has an effect on the latter. This effect is the 'intensified interest' (Brown & Levinson 1987) in

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

listening: being virtually addressed, it is in the recipients' intrinsic interest to attend to the reported utterance, to see "such things as: It's not yet complete, it's about to end, it just ended" (Sacks 1992: 649). That is, the pervasive use of constructed dialog in conversational storytelling is strategic: it serves the strategy of recipient design in that, via virtual participation and virtual turn-taking, the recipients' interest in listening, not only to the reported utterance, but, by extension, the story as a whole, is increased.[7]

In Section 5.3, the immediacy cline will be used as a diagnostic in the analysis of how report units are sequentialized at the level of utterance. It will be shown that in utterances with more than one report unit, narrators tend to order the units from less immediate to more immediate report units, thus sequentializing reports climactically.

Demonstration

A second crucial distinction intimately related to considerations of immediacy needs to be made between the two most immediate modes, Free Direct and Direct, and the remaining modes. Three distinctive features separate the two groups.

The first fundamental difference pertains to perspective, and, consequently, reference. First, as noted, Free Direct and Direct quotes use the reported speaker's referential expressions including tensed verbs, person deictics, and place deictics: reference is appropriate to the reported speaker, not the reporting speaker (see, for example, Holt 1996: 222, Coulmas 1986: 2, Mayes 1990: 346). The default 'ego-centric' organization of deictic expressions is lifted: deictics are transposed, or 'displaced' (Hanks 2011: 318), to some other origo, viz. "the person of the protagonist at the relevant time and place in a narrative" (Levinson 2004: 111; cf. also Hanks 1992: 56). The non-direct modes, conversely, use the reporting speaker's referential system: reference and origo are appropriate to the reporting speaker in the present reporting situation.[8] Second, given that Free Direct and Direct quotes are main clauses (Mayes 1990: 338) (whereas, for example, indirect presentations are realized in subordinate clauses), "direct quotes have fewer syntactic restrictions" (Mayes 1990: 338). They thus allow for the enquoting of non-clausal units, including most prominently vocatives, interjections, and exclamatives, which may be highly emotively charged and are typically not reportable in non-direct modes (see Banfield 1973: 7, Mayes 1990: 342–343). Third, only Free Direct and Direct enable reporters to imitate 'delivery aspects' (Clark & Gerrig 1990: 775) such as pausing, truncation, speech defects, inarticulate sounds, within-speech laughter, intonation, voice effects (pitch, quality, volume, drawling, etc.), and facial expressions and gestures (see Buchstaller 2002, Romaine & Lange 1991:240, Clark & Gerrig 1990).

All three characteristics (imitation of referential system, non-clausal units, and delivery aspects) concur such that in Free Direct and Direct quotation

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

the reporting speaker effectively 'theatricalizes' (Buchstaller 2002; see also Mayes 1990) the reported discourse: it is presented *as if* the reported speaker were present issuing the utterance him/herself in the present situation. Free Direct and Direct are hence *demonstrations* (Clark & Gerrig 1990: 802) while non-Direct reporting modes can be seen as *descriptions*. Seen as demonstrations, quotations using Free Direct and Direct mode are *mimetic*, facilitating the 'mimetic enactment' (Buchstaller 2002) of presented discourse, while all other modes are non-mimetic, or diagetic:

Demonstration	Description (Clark & Gerrig 1990)
Mimesis	**Diagesis** (Buchstaller 2002)
MDF MDD	MIF MII MUU MSS MVT MVV MRR MRQ

If we conceive of Free Direct and Direct reporting as mimetic and demonstrating, a crucial clarification pertains to the wide-spread assumption of quotation as verbatim reproduction. Clark & Gerrig (1990) have shown convincingly that the verbatim assumption is a fallacy (see also Tannen 1986, Mayes 1990, Thompson 1996). Conversationalists are neither cognitively able to reproduce discourse faithfully due to memory limitations (consider the impossibility to reproduce word by word an extended conversation one heard or participated in the day before) nor are they committed to such faithfulness.[9] In demonstration theory speakers are seen as "depict[ing] only selective aspects of what they are demonstrating" (Clark & Gerrig 1990: 799). The selective aspects speakers can choose from include first and foremost three main categories: delivery (vocal characteristics such as pitch, range, quality, gestures, emotional state, etc.), language (language proper, dialect, register, etc.), and linguistic act (illocutionary act, locutionary act, utterance act, etc.). Clark & Gerrig note that "[m]ost quotations depict illocutionary acts, including the propositions expressed, and treat the others as supportive or incidental" (Clark & Gerrig 1990: 779). In sum, in demonstration theory, the verbatim assumption is seen as a 'categorical error' (Clark & Gerrig 1990): uncommitted to verbatim faithfulness, speakers depict selected aspects only (for an attempt to reinstate the verbatim assumption and, thus, for a fundamental critique of demonstration theory, see Johnson & Lepore 2011).

If we, then, conceive of quoted speech as demonstrated and hence constructed speech, the limits to construction need to be acknowledged. One constraint to constructability is credibility: for the demonstration to be acceptable as a demonstration it needs to be credible. The demonstrated utterance should be recognizably utterance-like: it should 'sound' like naturally occurring speech. It appears reasonable to assume an 'utterance model,' a mental representation of the 'grammar' of utterances guiding speakers, to an extent, in designing quotation. Therefore, in investigating quotation, a yardstick against which the grammar of quotation can be examined is the utterance in turn-by-turn talk. If we accept the possibility of such

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

an utterance model, this does not preclude the possibility that the grammar of quotes diverges from the grammar of utterances. On the contrary, that divergence is highly likely, for credibility is not the only constraint to constructability. Another constraint is processibility: this principle demands "that the text should be presented in a manner which makes it easy for the hearer to decode in time" (Leech 1983: 64). As will be argued in more detail below, processing quotation, particularly multiple quotation (that is, sequences of quotes), poses specific problems in that reference in such quote sequences is in constant oscillation, thus exacerbating what I will call the boundary issue (see below): the difficulty for the listener to tell the voice of the quoting speaker from the voices of the quoted speakers. Therefore, the utterance model can only *co*-determine the grammar of quotation: a perfect utterance-quote fit is unlikely. In comparing quotes and utterances, a crucial interest thus lies in the *points of departure* from the utterance model, that is, the points where quotes significantly deviate from utterances. In Section 4.2, for example, it will be shown that interjections occur significantly more often at the onset of quotation than would be predicted on the basis of their occurrence in natural utterances, thus alleviating the 'boundary issue' discussed in the following subsection.

It is only superficially "something of a paradox" (Holt 1996: 226) that although quotation depicts rather than repeats an utterance, one of its key functions is as a non-grammaticalized evidential (Clift 2006), that is, as indexing the quoter's epistemic authority as a reliable source of information, a function which is significant in such diverse contexts as legal testimonies (Galatolo 2007), political discourse (Kuo 2001), nursing shift handover meetings (Bangerter *et al.* 2011), and even interactions between mediums and their audience (Wooffitt 2007). Although selective and depictive (and although participants in conversation are, probably, at least vaguely aware of this), quotation inevitably has "an air of objectivity" (Holt 1996: 236). This seeming objectivity is undoubtedly due to the three components of demonstration as a theatrical enactment outlined above: the switch into the reported speaker's referential system, the use of non-clausal units, and the imitation of delivery aspects. The quoted utterance has all the ingredients it takes to credibly uphold the pretense (see above) that the reported speaker issued the reported utterance 'in actual fact.'

The boundary issue

The fundamental difference between the demonstrative/mimetic modes Free Direct and Direct and the descriptive/diagetic modes Free Indirect, Indirect and so on has important repercussions: the mimetic/diagetic difference can be seen as a difference "in the speaker perspective or point of view of the reporter" (Coulmas 1986: 2). In the former, the "reporter steps back behind the characters whose words he purports to report" (Coulmas 1986: 2), in the

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

latter, "[h]e relates a speech event as he would relate any other event: from his own perspective" (ibid.). In other words: discourse presentation is not only a discoursal intertwining but, in the case of demonstration/quotation, also an intertwining of perspectives. I will argue in the following that the perspectival intertwining entails the necessity for the narrator to appropriately mark the boundaries of the perspectives adopted: I will refer to the issue as the boundary issue. The following excerpt illustrates the issue.

The narrator is a 57-year-old housewife, relating to her interlocutor how a friend's husband reacted to learning that their daughter had apparently had her first period:

(4.1) "Women problems" (Type: T30/Embed Level: EC1)

CNN		
CNI		
S1	PNS	[MVT Did I tell you about her little one who had stomach pains?] As she come back she said QSD [MDD Dad] [MDF what?] [MDF How long's our Mum going to be before she comes in?][MDF Another hour.] [MDF Oh.] [MDF Why?] [MDF Oh well I've got a bit of a stomach ache and I want to talk to her you know it's women problems.] [MDD All right] he said QSD. Well he knew QOO [MII what it was.] He said QSD [MDD you go up and lay in your bedroom] he said QSD [MDD and I can send her up when she comes in.] All right. The little 'un goes to bed. The little one's heard Mummy pull up on the drive and has come down the stairs [MVV well before anyone could say anything] [MSS he got it out.] [MDF She might have er period you'd better go and sort her out.] She said QSD [MDF what.] [MDF She might have one of her periods you'd better fucking go and sort her out. She's your daughter] She said QSD [MDD there's no need to [unclear] up and say it like that.] She said QSD [MDD you could have kept this shut and I could tell her myself.]
S1	PRR	Oh!
CNF		
S1	PNS	[MDD Well] he said QSD [MDD your fucking daughter you sort her out.]

(KBE-N2)

The narrator uses 21 instances of discourse presentation. Only the first (*Did I tell you about her little one who had stomach pains?*) refers to a discourse situation in which the presenter was involved; all other instances are attributed to speakers not present in the telling situation. The narrator is animating the speech of three non-present speakers: her friend's husband's, her

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

friend's, and their daughter's. The narrator uses a restricted set of presentation types. As is shown by the tags, eight utterances are presented using Direct (MDD), nine using Free Direct (MDF); one Speech Act Representation (MSS) (*He got it out*) and one Indirect presentation (MII) (*Well he knew what it was*) are used. Discourse presentation in the excerpt is thus predominantly (Free) Direct.

In terms of perspective, the narrator is switching to and fro between no less than four different perspectives: the three characters' perspectives and her own perspective (expressed for example in the preface *Did I tell you* . . ., narrative talk, e.g., *As she come back* . . ., and all reporting clauses *he said*, *she said*, etc.). The switches in perspective can be understood in terms of Goffman's (1981) production format as switches from 'author,' who is "formulating his own text" (1981: 145), to 'animator,' that is, "the talking machine, a body engaged in acoustic activity" (1981: 144) animating someone else's text. The perspectival and role switches performed in the main part of the narrative are represented in (4.2); the three characters are referred to as 'father,' 'mother,' and 'daughter':

(4.2)

Perspective	Role	Realization
Presenter	*Author*	[^{MVT} Did I tell you about her little one who had stomach pains?] As she come back she said ^{QSD}
Daughter	*Animator*	[^{MDD}Dad]
Father	*Animator*	[^{MDF}what?]
Daughter	*Animator*	[^{MDF}How long's our Mum going to be before she comes in?]
Father	*Animator*	[^{MDF}Another hour.]
Daughter	*Animator*	[^{MDF}Oh.]
Father	*Animator*	[^{MDF}Why?]
Daughter	*Animator*	[^{MDF}Oh well I've got a bit of a stomach ache and I want to talk to her you know it's women problems.]
Father	*Animator*	[^{MDD} All right]
Father	*Author*	he said^{QSD}. Well he knew ^{QOO} [^{MII}what it was.] He said ^{QSD}

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

Father	*Animator*	[^{MDD}you go up and lay in your bedroom]
Presenter	*Author*	he said ^{QSD}
Father	*Animator*	[^{MDD}and I can send her up when she comes in.]
Presenter	*Author*	All right. The little 'un goes to bed. The little one's heard Mummy pull up on the drive and has come down the stairs [^{MVV}well before anyone could say anything] [^{MSS}he got it out.]
Father	*Animator*	[^{MDF}She might have er period you'd better go and sort her out.]
Presenter	*Author*	She said ^{QSD}
Mother	*Animator*	[^{MDF}what.]
Father	*Animator*	[^{MDF}She might have one of her periods you'd better fucking go and sort her out. She's your daughter]
Presenter	*Animator*	She said ^{QSD}
Mother	*Animator*	[^{MDD}there's no need to [unclear] up and say it like that.]
Presenter	*Author*	She said ^{QSD}
Mother	*Animator*	[^{MDD}you could have kept this shut and I could tell her myself.]

In the excerpt presented in (4.2), the presenter constantly shuttles between distinct roles (animator and author) and perspectives (her own, the father's, the mother's, and the daughter's). The narrator's discourse is highly polyperspectival: altogether, 22 role and perspectival shifts occur. Crucially, the shifts in role and perspective entail shifts in reference: the numerous occurrences of referring pronouns *I/me/myself*, *you/your*, and *she/her* oscillate in reference across the whole narrative (only the male pronoun tokens referring to the only male character involved do not undergo this shift in reference). This oscillation in reference poses heavy demands on the listener: each time the perspective shifts reference must be reassigned to changed referents. The crux in reassigning reference is the question *when* to perform the reassignment. The reference encoded in referring expressions is not decoded for each expression individually but in the context of the discourse unit in which the

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

expression is used. In conversation, the basic unit is the utterance: principally, the references made in a speaker's utterance are resolved in accordance with the same speaker's referential system: his/her *I*s and *you*s are processed as referring to him/herself and, respectively, his/her immediate interlocutor(s), respectively. In the case of quotation, the underlying unit is the report unit, that is, the locutionary/phatic act the presenter is animating. In other words, to perform reference reassignment successfully it is imperative for the recipient to be able to discern the boundaries of quotation: where the quote starts and where it ends. The issue at stake is then the boundary issue. In writing, readers are helped in the task of reassigning reference by means of quotation marks, which unambiguously demarcate the end and the beginning of quotes. In conversation, as a visual and auditory medium, quotations need to be marked using visual and/or auditory means. Visual markers, such as, potentially, facial expression or gaze, cannot be studied in the NC (as in most other corpora). Apart from obvious lexical quotation markers such as quotatives, which have seen a surge in interest with the emergence of newcomers such as GO and BE *like*, quotation marking has attracted little systematic, let alone quantitative analysis so far. For example, Holt (1996: 223) notes, somewhat in passing, that "changes in prosody can be used to mark a shift from unreported speech to DRS [direct reported speech]" (see also Tannen 1986: 320). The problem of the 'unquote,' that is, how the exit from quotation is marked, has largely gone unnoticed, at least for English; it has, however, been noted by researchers working on languages other than English, for example, Russian (Bolden 2004) and German (Golato 2000).

In the main part of this chapter I will be concerned with two broad types of such means narrators use to assist recipients in attending to the boundary issue: linguistic and paralinguistic. The linguistic means used by narrators to mark quotation I will investigate include interjections (Section 4.2), while the paralinguistic means investigated include silent pauses (Sections 4.3).

4.2 Interjections as quotation markers

4.2.1 Introduction

This section reports on a case study on the positioning of interjections in quotes in narrative utterances. Given that, as noted, the conversational utterance can be seen as the reference model for quotation, interjections in quotes will be contrasted with interjections used in utterances in non-narrative conversation. Specifically, I will be concerned with interjections in 'first position.' The notion of first position refers strictly to the position of the immediately first word. First position is thus distinguished from more 'liberal' notions such as 'initial position', which may be used to refer to text-early positional *segments*. Drawing on Hoey's (2005) theory of lexical priming, I will argue that interjections in quotations acquire, in addition to the variety of

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

roles they fulfill in normal conversation, a new function that is not found in non-quotative discourse: a discourse-deictic function of marking quotation. I will discuss the use of interjections as quotation markers in keeping with the recipient-design principle: as a service to the recipients by which narrators flag the imminence of quotation, the switch in reference system that is effected by (Free) Direct quotation, and the need for the recipients to resolve reference polyperspectivally, that is, as according to the perspectives of the various reported speakers.

Before embarking on the case study, let us explore the terrain: what are interjections? A useful starting point definition is as members of the class of inserts, which Biber *et al.* (1999: 56) describe thus:

> They do not form an integral part of syntactic structure, but are inserted rather freely in the text. They are often marked off by intonation, pauses, or by punctuation marks in writing. They characteristically carry emotional and interactional meaning and are especially frequent in spoken texts.

That is, four features typify inserts and hence also interjections: syntactic isolation, intonational separation, emotional and interactional meaning, and virtual restriction to speech. Given that the NC is a speech-only corpus and not annotated for intonation, only syntactic isolation and the expressive and interactional significance of inserts are of special relevance in the present study. Particularly distinctive is their 'compositional' agility: because inserts are syntactically free elements which normally do not enter into constructions with other words, they tend to occur either utterance-initially or constitute utterances by themselves (see Ameka 1992: 105). This sets them apart from other typically utterance-initial items such as the pronoun *I* and the coordinator *and*, which, under normal circumstances, cannot constitute complete utterances.[10] Inserts come in a wide variety: besides falling into the classes of greetings, discourse markers, attention signals, hesitators and other subtypes (see Biber *et al.* 1999: 9394), inserts include first and foremost interjections.

Ameka (1992) distinguishes two basic types of interjections: primary and secondary. Primary interjections, such as *oh* and *cor*, are defined as "relatively conventionalised vocal gestures (or more generally, linguistic gestures) which express a speaker's mental state, action or attitude or reaction to a situation" (Ameka 1992: 106). Contrary to 'secondary interjections' such as *Holy Christ*, *shit*, etc., which are forms that belong to other word classes and can thus be used in non-interjectional functions, primary interjections are "not used otherwise" (Ameka 1992: 105). Interjections pose serious problems for classification: they can serve as backchannels (*mm*), hesitation markers (*erm*), attention signals (*hey*), expletives (*god*), ad hoc vocalizations (*whoa*), and discourse markers (*yeah*) (see Quirk *et al.* 1985, Norrick forthcoming). A functional classification of interjections is proposed in Ameka based on the functions of language developed by Bühler (1990 [1934]) and Jakobson (1960).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

Three functional types of interjections are distinguished. First, the 'expressive' function, performed by interjections which are "symptoms of the speaker's mental state" (Ameka 1992: 113). Two subfunctions of the expressive function include the 'emotive,' which expresses "the speaker's state with regard to the emotions and sensations they have at the time" (Ameka 1992: 113), and the 'cognitive,' which indexes "the speaker's state of knowledge and thoughts at the time of utterance" (Ameka 1992: 113). An example of the emotive function is *urgh*, expressing disgust, an example of cognitive interjections is *ah* as a response to new and/or surprising information (see Aijmer 1987). Second, interjections can have a conative function, "aimed at getting the hearer's attention or demanding an action or response" (Ameka 1992: 113). An example is *sh!*, which signals that the speaker wishes the hearer to be quiet. Third, interjections can be 'phatic'; their primary purpose is "the establishment and maintenance of communicative contact" (ibid.). Examples include *hm* and *mhm*. Therefore, the common view that interjections only or even mainly "have an exclamatory function, expressive of the speaker's emotion" (Biber *et al.* 1999: 1083) falls short of reflecting this functional breadth (see also Norrick 2009: 868).

Another defining feature of interjections seems to be their tendency to occur utterance-initially. While this tendency has been noted repeatedly (see, for example, Norrick 2008, 2009, forthcoming) it has been less documented quantitatively. Below, I will provide evidence to validate the notion.

Further, interjections have been observed to co-occur with quotation. Svartvik (1980), for example, mentions, somewhat in passing, the association of *well* with quotatives and the onset of quotation. Functionally, he interprets this use of *well* as "a signal indicating the beginning of direct speech, parallel to that of quotation marks in writing" (Svartvik 1980: 175). Schiffrin (1985) too noted the co-occurrence of *well* with quotation; specifically, she discovered a preference for *well* to mark "reported responses" (Schiffrin 1985: 658). Jucker (1993) found that when used as a 'frame' serving "to separate discourse units" (1993: 446), *well* often introduces direct speech. Aijmer (1987) observed that *oh* and *ah* co-occur with quotation particularly where a change in voice (accent, volume, speed, etc.) was involved. Holt (1996: 236) discovered that "*well* and *oh* are common at the start of DRS [direct reported speech]" noting also, in passing, that *well* and *oh* at the onset of quotation "help to indicate to the recipient that the speaker is now quoting and that the utterance should not be interpreted as the speaker's but as the reported speaker's." Biber *et al.* (1999) dedicate a small section to what they term 'utterance openers,' defined as items by which speakers "signal that they are embarking on direct speech quotation" (Biber *et al.* 1999: 1118), a functional analysis clearly reminiscent of Svartvik's quoted above. The items identified by Biber *et al.* include *oh*, *well*, *look*, and *okay*. Both Svartvik, Schiffrin, Aijmer, Holt, and Biber *et al.* are qualitative studies: the findings are not quantified. A more quantitatively oriented (but in terms of quantification still preliminary) case study is

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

Rühlemann (2007: 139ff.) who reports on frequencies of *oh*, *well*, and *look* as collocates of the quotatives SAY, THINK, and GO. Crucially, all these studies converge in noting that interjections used as signals of quotation occur quotation-initially.

What all of the above-mentioned studies fail to provide is evidence to support the notion that the function of interjections as quotation markers *differs* in any way from the usual functions of interjections in utterances. If interjections occur utterance-initially *and* quote-initially, it does not automatically follow that the occurrence in initial position in constructed dialog signifies a (new) function as a 'signal' that speakers 'are embarking on direct speech quotation.' It could be simply that the way interjections occur in direct reports is a faithful mimesis of the way they occur in natural utterances. For example, McCarthy (1997) notes that *oh* and *well* are often included in quotation "in the way they might be expected to occur in real conversation" (McCarthy 1997: 159). If that were indeed the case, no special signal, or marking, function could be ascribed to them. In order to establish whether or not interjections in quotation gain an additional discourse-deictic marking function as auditory quotation marks one will have to compare the frequencies of occurrence of interjections in (i) (Free) Direct reports and (ii) conversational utterances. Only if a significant overuse in quotes can be observed will it be possible to claim such a function as a marker of quotation.

In the following case study on interjections, precisely this comparison between interjections as quote- and utterance-first items will be performed. The identification of interjections will be largely based on the POS tagging available in the NC. One immediate limitation to this methodology needs to be acknowledged: interjections are "not consistently marked in corpora" (Norrick forthcoming). The BNC and, by consequence, the NC are no exception to this: the Claws5 (c5) tagset recognizes *primary* interjections only. There is thus no way of automatically identifying and analyzing secondary interjections such as *oh my god* (components of) which can be used non-interjectionally. While it seems that most secondary interjections are much less frequent than primary interjections (see Norrick forthcoming), there is one secondary interjection of outstandingly high frequency: the item *well*, which is, in conversation, predominantly used as a discourse marker, particularly in first position. This item is consistently annotated as ADV (adverb) in the BNC and the NC. Given its high frequency and prominent role in first position, it will be included in the analysis.[11]

4.2.2 Data and methods

It is useful to reiterate what I refer to as 'first position.' Before, I used the metaphorical expression of 'auditory quotation marks' to describe the (assumed) role of interjections as markers of quotation. The metaphor is only partially appropriate. Quotation marks in writing mark both the left and

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

the right boundary of quotation; all previous research on ITJ in quotation, however, points to interjections occurring in the left-boundary slot only, that is, as the very first word in the quote: it is this position that I refer to as first position. So the present analysis will target not interjections in any position but specifically in that first position defined as the immediately first word, both in conversational (i.e., non-narrative) utterances and (Free) Direct quotes (occurring within narrative utterances). A second important constraint relates to the notion of 'first': strictly speaking, an item can only be considered first if it is followed by some material, a second word, a third and so on; an item that is free-standing, is, in this sense, not first although it occurs in the left-boundary slot. That is, the analysis will exclude 'lone' interjections not preceding anything within the same unit (quote or utterance). The minimal requirement for the quotes and utterances I am going to look at is thus that the number of words they enclose be $n > 1$.

The hypotheses guiding the following analysis are:

H_0: Certain interjections do not occur in first position more frequently in quotes ($n > 1$) than in conversational utterances ($n > 1$).
H_1: Certain interjections occur in first position more frequently in quotes ($n > 1$) than in conversational utterances ($n > 1$).

How are positional data obtained? Positions can be defined as the proportions of the words preceding the item in question out of all words in a unit (see, for example, Rühlemann *et al.* 2011). Positional data for interjections in quotes and utterances are thus obtained by dividing the number of words preceding the interjection by the total number of words in the quote/utterance. Thus, positional values normally range between 0 and < 1 (but see Section 4.3).[12] For example, in (4.3), a quote with a total of 13 words, the positions of the two instances of *ooh* at the beginning and the end are obtained from the two divisions 0 / 13 = 0 and 12 / 13 = 0.92:

(4.3) [^{MDD}**ooh** this is good value for [laughing] four ninety nine! Int it leather? **Ooh**!] (KDV-N2)

Using XQuery, the positions of interjections tagged ITJ were calculated for two data sets: (i) for conversational utterances occurring in the pre-narrative and, respectively, post-narrative components CPR and CPO, and (ii) for quotes (i.e., Free Direct and Direct reports) occurring within the narrative utterances in the narrative components (CNN). The queries also targeted *well*. Incorrect matches, namely instances of *better*, whose headword is *well* too, were removed. The correct matches were integrated into the matches for the queries for ITJ.[13]

The data under scrutiny are sizable: there are more than 3,000 interjections in conversational utterances and roughly 600 interjections in quotes longer than 1 word.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

Table 4.1 *Number of interjections (ITJ), including* well, *in quotes and conversational utterances longer than one word (n > 1)*

Utterances ($n_{Words} > 1$)	Quotes ($n_{Words} > 1$)
3,007	597

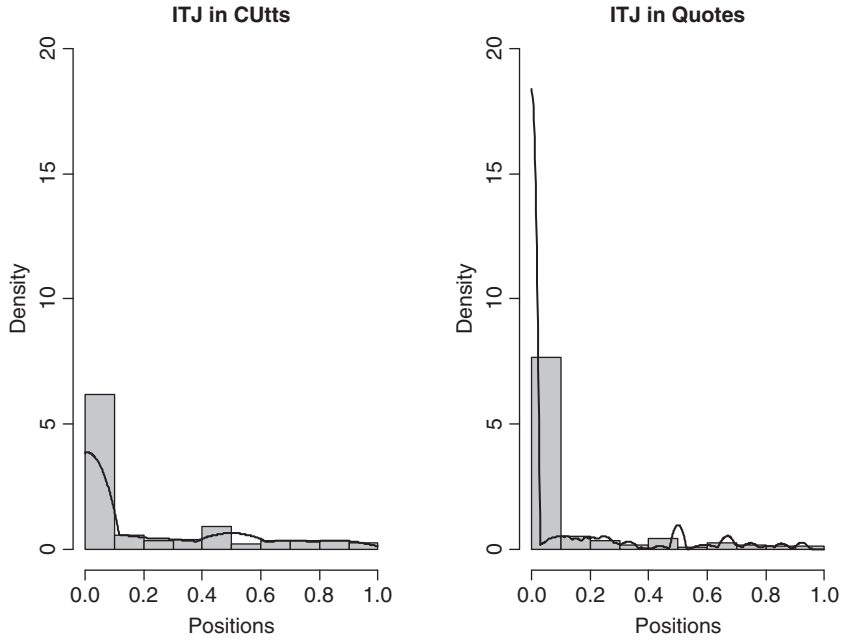


Figure 4.3 Histograms with density curves of positions of interjections (ITJ), including *well*, in conversational utterances (CUtts) and quotes longer than one word

4.2.3 Results

To initially approach the data, consider the histograms in Figure 4.3, representing the distribution of positions of interjections in the two subsets.

The density curves in the histograms in Figure 4.3 represent the probabilities of occurrence of positions of interjections estimated on the basis of the positional frequencies (see Dalgaard 2008: 59). The histogram in the left panel provides some initial evidence to support the above cited observation that interjections tend to occur early in utterances: the first bin, representing the positional values between 0 and 0.1, is clearly the tallest, and the density curve has a very clear maximum in this segment. However, the first bin in the right panel, representing positions of interjections in *quotes*, is even taller than

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

Table 4.2 *10 most frequent ITJ in conversational utterances (CUtts)*

N	ITJ	Freq_total	Freq_pos=0	%_pos=0
1	yeah	623	357	57
2	well	587	300	51
3	oh	585	430	74
4	no	440	246	56
5	mm	203	105	52
6	yes	164	86	52
7	ah	76	46	61
8	ooh	40	20	50
9	aye	37	23	62
10	ha	28	4	14
	Total	2,783	1,617	58

that for utterances. What is more, the height of the curve's primary maximum for interjections in quotes is many times the height of the curve for interjections in utterances. That is, the relative probabilities of finding interjections in very early positions in quotes are dramatically higher than to find them in early positions in utterances. Note also that in both panels the peak is precisely at positional value 0 for first position. This can be taken as preliminary evidence that interjections both in utterances and quotes may be *primed* to occur in that first position. What we do not know yet though is whether this is a general tendency applying to all or most interjections or whether the steep peaks are 'powered' by only a few interjections and can thus not unreservedly be generalized. It is therefore necessary to establish, not how the *class* of interjections, but how individual interjections behave positionally.[14]

In the following, I identify the most frequent interjections in utterances and quotes as well as their frequencies in first position. I begin with interjections in conversational utterances.

Table 4.2 lists the top ten most frequent interjections in conversational utterances; it also gives the frequencies of occurrence in position 0 (that is, as the very first word).

As can be seen in Table 4.2, the three most frequent interjections in conversational utterances in the NC are *yeah*, *well*, and *oh* – a good match of the top three interjections in large general corpora such as the Longman Spoken and Written English Corpus (LSWEC), where, in that order, *yeah*, *oh*, and *well* (see Norrick 2008: 444) were the three most frequent interjections. As regards positioning, *oh* has both the highest absolute frequency for first position (430 occurrences) and the highest proportion for that position (74%). Generally, it can be seen that, for nine out of the ten most frequent items, the proportion of occurrence in first position (where the positional value is 0) is above 50% (only *ha* has a (much) lower percentage), while the overall proportion for occurrence in first position for all ten items is 58%. Note that the nine interjections with high proportions have a combined

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

frequency of occurrence of 2,727 (out of 3,007) and thus account for 91% of all the interjections in conversational utterances. This is clear evidence to suggest that the above-noted trend for interjections to overall occur in utterance-first position is indeed a reflection of a *general* trend which is powered not by a few isolated interjections but by almost all of the most frequent interjections.

So far, we have focussed on frequency of occurrence in first position. How are interjections in conversational utterances distributed across all positions?

Taking the median, indicated in the bold vertical lines in the boxes in Figure 4.4, as a measure of central tendency, we see that only the positions of *ooh* and *ha* have a median value greater than 0. The medians for the remaining eight interjections are invariably 0. This is further evidence to suggest that the trend for the most frequent interjections to occur in first position is indeed a central trend. Also, the outstanding status of *oh* is confirmed: it is not only, as noted, the item with the highest share in occurrence in first position, but its IQR (the interquartile range comprising the 50% of the data around

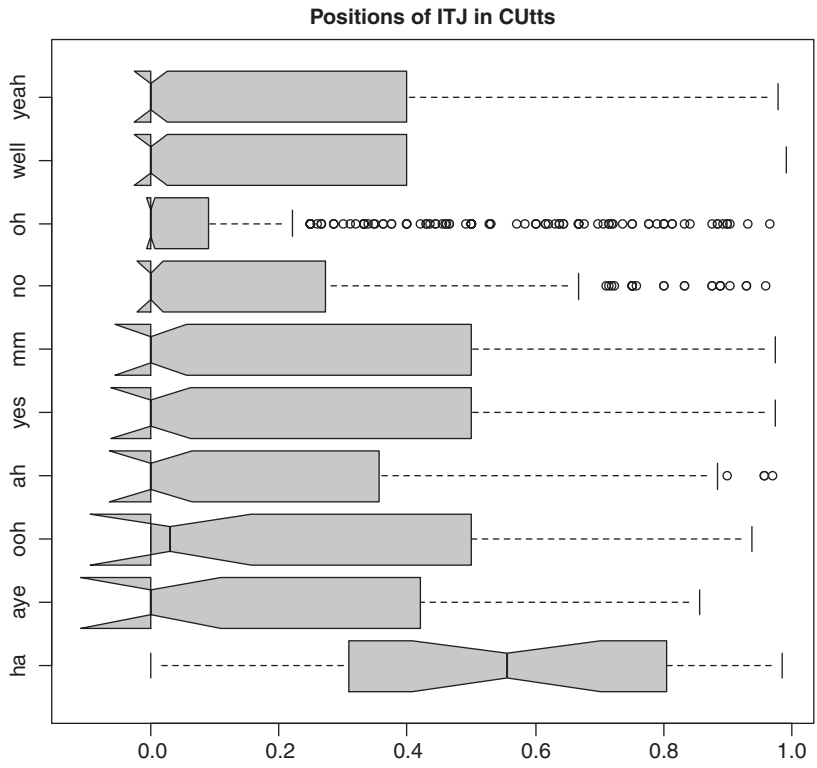


Figure 4.4 Boxplots of positions of the ten most frequent interjections in conversational utterances (CUtts)

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

the median) is by far the smallest. This indicates that its positional homogeneity is greatest. Given this homogeneity and its high frequency of occurrence, the notch in the box for *oh* (depicting the interval in which the true median is estimated to lie) is minimal. This suggests we can be highly confident that the true median is indeed 0. The notches for *yeah*, *well*, and *no* are slightly wider suggesting we can be slightly less confident that the true median will be 0 too. The notches for the remaining items, however, are much wider, warranting much less confidence. As regards dispersion, the second least dispersed positions (after those for *oh*) are for *no*, where the IQR is 0.27, while the IQRs for *yeah* and *well* are both 0.4. Finally, the IQR for all eight interjections with median position 0 is 0.33, that is, exactly within the first third of the positional range between 0 and 1. In sum, we can say that interjections in conversational utterances have a very clear tendency to occur not randomly across positions but within the first third of the utterance and, particularly, at the very onset of the utterance (that is, in first position). This can be said with great confidence of the interjections *oh* and, with slightly less but still sufficient confidence, of the items *well*, *yeah*, and *no*. In priming terms: *oh* and, to a lesser degree, *yeah*, *well*, and *no*, are primed for use not only initially but also, more specifically, for use in first position in conversational utterances.

How are positions of interjections distributed in *quotes*?

First, note that the list presented in Table 4.3 only partially overlaps with the list of items identified in previous research (see above), which included *oh*, *well*, *ah*, *look*, and *okay*. The forms *look* and *okay* mentioned in Biber *et al.* (1999) are notably missing from Table 4.3. This is unsurprising since they are not tagged as interjections. However, even lexical searches for positions of *look* and *okay* in quotes returned very low figures: *look* occurred 7 times and *okay* just once in quote-first position. I suspect that the items may be more frequent in quote-first position in AmE, the variety characteristic of a large part of Biber *et al.*'s corpus.

Table 4.3 *10 most frequent ITJ in quotes*

N	ITJ	Freq_total	Freq_pos=0	%_pos=0
1	oh	186	168	90
2	well	174	131	75
3	no	54	37	69
4	yeah	37	21	57
5	ooh	23	20	87
6	yes	23	17	74
7	hello	10	8	80
8	ah	7	7	100
9	ha	15	4	27
10	blah	10	2	20
	Total	539	415	77

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

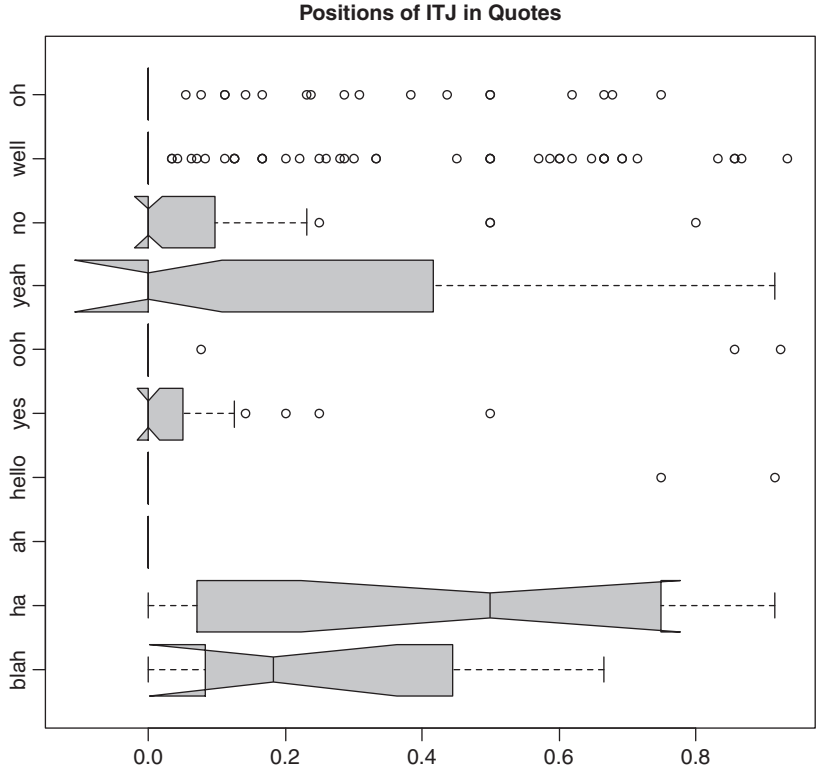


Figure 4.5 Boxplots of positions of the ten most frequent interjections in quotes

The following differences will be noted if we compare the list of the ten most frequent interjections in quotes shown in Table 4.3 with those most frequent in conversational utterances shown above in Figure 4.5. First, *mm* and *aye* are missing from the list for quotes, which instead contains *hello* and *blah*. While the frequency of *aye* in utterances is comparatively low, the frequency of *mm* in utterances is not. Its absence from the list of interjections in quotes is therefore curious and worthy of further investigation.[15] Second, the proportions of occurrence in first position seem to be higher in quotes than in utterances, both overall (note the overall proportion of 77% compared to 58% for utterances) and for specific items: these include most notably *ah*, where the proportion of use in first position is 100% for quotes compared to 61% for utterances, *oh* (90% vs. 74%), *well* (75% vs. 51%) , and *no* (69% vs. 56%). The proportions of use in first position for *yeah*, by contrast, are the same (57%).

The boxplots in Figure 4.5 show that, not surprisingly, the notches (indicating the 95% confidence interval) are extremely wide for the less frequent items such as *blah* and *ha*, which we can therefore safely disregard.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

Table 4.4 *Chi-squared statistics for* oh *and* well *in quote- and utterance-first position (df = 1)*

	X-squared	p-value	Cramer's V
oh	22.1974	2.46E-06	0.197108
well	30.6227	3.134e-08	0.2003363

However, the notch for *yeah* is also wide so that, for *yeah*, it cannot be said with confidence that, in quotes, its true positional median is 0. That this true median is 0 can be maintained with great confidence for *oh* and *well*: for these two items the attraction to, or priming for, first position is so dramatic that the interquartile range (IQR) *coincides* with that median, which is why no boxes are drawn around the median (*ooh*, *hello*, and *ah*, for which no boxes are drawn either, are of much lesser frequency than *oh* and *well* and no priming claims will therefore be made for them). Accordingly, the IQRs for *oh* and *well* are 0, indicating that the homogeneity of their distribution could hardly be greater. These two items, then, exhibit a very strong attraction to first position in quotes. In priming terms: *oh* and *well* are primed for use as the immediately first word in quotes. As regards the remaining eight interjections, their frequencies of occurrence in quotes are below 100 and hence too low to warrant strong priming claims. These items are therefore not investigated further.[16]

To wrap up the findings on *oh* and *well* so far, the analysis discovered evidence that both *oh* and *well* are strongly attracted to first position both in conversational utterances and in quotes. It was found that this attraction to first position is greater in quotes than in utterances. To test this latter finding for significance, chi-squared tests were carried out. The results are summarized in Table 4.4.

We see that the associations for *oh* and *well* are very highly significant: the p-values for the two items are by far smaller than 0.05. The effect sizes, measured by the Cramer's V values, can be considered modestly strong: the values are roughly 0.2 for both items (Cramer's V ranges between 0 and 1). The H_0, which stated that certain interjections occur as frequently in first position in quotes as in conversational utterances, can hence be rejected: *oh* and *well* are used significantly more often in quote-first than utterance-first position. *Oh* and *well*, then, are precisely not included in quotes 'in the way they might be expected to occur in real conversation' (see McCarthy 1997: 159), they are included in quote-first position much more often than might be expected on the basis of their occurrence in utterance-first position.

4.2.4 Discussion

What to make of this significant result? At first, I briefly review previous research on *oh* and *well*.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

Well and, to a lesser extent, *oh* have been extensively researched. *Oh* can be classed as a primary interjection, serving a cognitive function, namely flagging the speaker's surprise on receiving new, maybe unexpected information: "its core function appears to be to convey some degree of surprise, unexpectedness, or emotive arousal" (Biber *et al.* 1999: 1083). Schiffrin (1987: 73) downplays the emotive potential of *oh* and stresses its role as a marker of information management: "*oh* pulls from the flow of information in discourse a temporary focus of attention which is the target of self and/or other management." Similarly, Heritage (1998), investigating the use of *oh* in responses to questions, finds *oh*-prefaced responses "recurrently associated with inquiries that address 'known information' or other difficulties concerning the appositeness or relevance of a question." Aijmer (1987) notes its special importance due to its high frequency, which places *oh* "among the most frequent words" in her corpus (the London-Lund Corpus). Additionally, she interprets *oh*, and also *ah*, which she sees as largely comparable to *oh*, in the light of relevance theory (see Sperber & Wilson 1986) as devices a speaker uses to "indicate that he takes the previous utterance as relevant" (Aijmer 1987: 84). In Aijmer (1996: 221), finally, *oh* is seen as a discourse deictic "pointing only backwards," that is, as framing *preceding* information as new, surprising etc. (see also Heritage 1998).

Well, on the other hand, is only rarely treated as an interjection (e.g., Norrick 2008); more commonly, it is seen as a discourse marker. The common denominator to descriptions of how *well* marks discourse seems to be the assumption that *well* "introduces a part of the discourse that has something in common with what went before but also differs from it to some degree" (Svartvik 1980). In a similar vein, Schegloff & Lerner (2009: 91) interpret *well*-prefaced responses "as general alerts that indicate nonstraightforwardness in responding." More narrowly, Fraser views *well* as a marker of 'dissonance' (Fraser 1990: 387); Schiffrin argues that *well* introduces "a temporary suspension ... for immediate coherence of a response" (Schiffrin 1985: 648) and is therefore "likely to be used when respondents cannot easily meet conversational demands for response" (Schiffrin 1985: 650; see also Schiffrin 1987: 323). Levinson (1983: 334) notes that *well* "standardly prefaces and marks dispreferreds" (that is, dispreferred second-pair parts in adjacency pairs). Jucker (1993: 435), finally, adopting a relevance-theoretic viewpoint, argues that *well* is used to signal that "the most immediately accessible context is not the most relevant one for the interpretation of the impending utterance." Importantly, unlike *oh*, *well* is not only backward-looking as prefacing *second*-pair parts but also forward-looking in orientation as it announces the (partial) non-compliance with conversational expectations.

Bearing these functional interpretations of *oh* and *well* in mind, two possible interpretations of the significant results for *oh* and *well* at the onset of quotes come to mind: (i) narrators quote more utterances which express, broadly speaking, surprise and deviate from expectations than there are in

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

actual conversation and (ii) narrators use *oh* and *well* as auditory quotation markers alleviating the boundary problem of quotation at the left periphery. I will discuss the two interpretations in what follows.[17]

Upon closer scrutiny, interpretation (i) is less implausible than it may appear at first sight. First, demonstration theory holds that speakers depict utterances selectively (see Section 4.1). If selection is thus a design principle in using quotation, it cannot be ruled out that speakers can also *pre*-select certain utterance types *for* presentation. Given the two interjections' core functions as indices of surprise (*oh*) and dispreferreds (*well*), the pre-selection of *oh*- and *well*-introduced utterances for quotation would amount to a pre-selection of utterances expressing surprise and deviation from expectations. Such utterances might be perceived as intrinsically interesting to the recipient, loosely speaking, as 'more dramatic.' Their pre-selection could thus serve to intensify the recipient's interest in the storytelling (see Brown & Levinson 1987). In other words, pre-selection of *oh*- and *well*-prefaced utterances for quotation would be a strategy in the service of recipient design.

Second, to support this theory it may be argued that if indeed *oh* and *well* acquired a discourse-deictic function as quotation markers this new function would possibly have an effect upon the usual functions *oh* and *well* carry out in (non-quoted) utterances. It might be the case that when used to flag quotation the items will no longer fulfill the usual utterance functions such as to express surprise, acknowledge receipt of new information, and preface deviation from expectations and so on and instead only attend to the new task of marking quotation. That deletion effect, however, does not seem to exist. It is easy to find in the data examples of *oh* and *well* used at the onset of quotation that still preserve their 'original' meanings. Consider (4.4) to (4.7), which illustrate uses of *oh* and *well* in quote-first position.

oh:

(4.4) black welding glasses on, and he turned round and he made me jump. Like^{QLK} [^{MDD} **oh**, Colin,] and
(KB1-N2)

(4.5) [^{MDD}I locked all [unclear] off from under my windows] I says^{QSZ} .
[^{MDD}And I'm about six foot four and I'll handle anyone that breaks in here.] [^{MDD} **Oh** that's alright then] he says^{QSZ}
(KDY-N1)

well:

(4.6) she said^{QSD} [^{MDD}it was all packed up in a box] and I said^{QSD} [^{MDD} **well** you wouldn't see it then would you?]
(KE6-N2)

(4.7) Women have babies. I said ^{QSD} [^{MDD} **well** it comes from a man.] So Luke said ^{QSD} [^{MDD}**well** what are they like?] So I said ^{QSD} [^{MDD}**well** they're like little tadpoles.] And he went ^{QGD} [^{MDD} **well** I can't feel any.]
(KC5-N1)

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

In (4.4), the presenter reports the speaker's surprise (*he made me jump*) at recognizing Colin (*oh Colin*) 'hidden' behind black welding glasses. In (4.5), the quote *Oh that's alright then* acknowledges the new information conveyed by the interlocutor's description of his tall build (*I'm about six foot four*) and his reassurance that he'll *handle anyone that breaks in*. In (4.6), the female interlocutor is apparently looking for something *although* she had it *all packed up in a box*, to which the quoted reply is a hint that she might not be able to find it *because* she packed it in a box; the reply then carries a critical overtone thus making it a dispreferred. (4.7), finally, contains quoted efforts by adults to explain to a boy what sperms are. The explanations given are evasive (*well it comes from a man*) and iconic (*well they're like little tadpoles*), with *well* at the onset signalling the adults' uncertainty as to whether their descriptions will be adequate for the child. The quoted responses by the boy also marked by *well* seem indeed to contain a degree of dissatisfaction: in response to the adults' vague description *it comes from a man* the boy demands a more specific description (*well what are they like?*) and the adults' tadpole simile is overtly rejected (*well I can't feel any*). Note that these are randomly selected examples and that more such examples could probably be found. Upon inspection of the data it appears that *oh* and *well* keep their usual discourse-marking functions even in quotation. The assumed additional quotation-marking function does not cancel them out or noticeably compromise them. This, then, the argument concludes, could be evidence that *oh* and *well* have no additional quotation-marking function.

How to counter these arguments? First, one weakness of the selectivity argument is that there is no evidence to support one important entailment. This entailment is that if we assume that speakers select utterances conveying surprise and dispreferreds for their greater intrinsic interest for the recipient we would in a similar vein (have to) expect that speakers *generally* select more frequently for presentation those utterances that 'score' higher in terms of dramatic potential, including, for example, utterances conveying strong emotions such as disgust, anger, amazement, or pain. I am not aware of any evidence (in any research) supporting this notion. Nor does the NC provide such evidence. If interjections are taken as possible indicators for these kinds of utterances, searches for interjections commonly associated with these emotions in quote-first and, respectively, utterance-first position do not uncover evidence to support the notion. As can be seen in Table 4.5, candidate interjections such as *urgh* and *ugh* for disgust, *fuck* for anger, *wow* for amazement, and *aargh* or *ow* for pain, hardly ever occur either in quote- or utterance-first position:

To be sure, utterances expressing strong emotions can be made in many other ways. It can therefore not be ruled out that indeed the selection-by-dramatic-potential is a principle that applies generally rather than only to utterances conveying surprise and dispreferreds. However, its general

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

Table 4.5 *Frequencies of* fuck, wow, urgh/ugh, *and* aargh/ow *in quote- and utterance-first position*

ITJ	Freq. quote-first	Freq. utterance-first
fuck	0	1
wow	0	1
urgh/ugh	2	1
aargh/ow	0	1

application needs yet to be demonstrated. Given the current unavailability of direct evidence, the notion cannot unreservedly be adopted.

To counter the second argument brought forth above, that if *oh* and *well* acted indeed as quotation markers their usual discourse-marking functions would likely have to disappear or at least lose in force, it can be said that this argument rests on the assumption that there is a one-to-one relationship between item and function: any item can fulfill only any one function. It is widely acknowledged that high-frequency features (such as *oh* and *well*) are highly multi-functional in general (see Schiffrin 1987) and that discourse markers are particularly prone to carrying out multiple functions in different contexts. Redeker (1991), for example, defines discourse markers as items "uttered with the *primary* function of bringing to the listener's attention a particular kind of linkage of the upcoming utterance with the immediate discourse context" (1991: 1168; added emphasis). The notion of 'primary function' implies the existence of secondary functions. Thus it is useful to distinguish different levels of functionality. If we then assume a quotation marking role for *oh* and *well* and if we further bear in mind that, as shown above, the usual discourse-marking functions as well as their directionalities do remain intact, the assumed quotation-marker function could be seen as a function performed and processed at an additional higher level of functionality. This functional layering can be represented thus, using quotes from examples (4.5) and (4.7) for illustration:

Marking	Directionality	Marking	Directionality
quotation	forward	quotation	forward
surprise	backward	dispreferred	bi-directional
Oh that's alright then		well I can't feel any	

In *Oh that's alright then*, *oh* fulfills a dual role. It registers the reported speaker's reception of new, potentially surprising information; in this function *oh*'s directionality is backward-looking. At the same time, due to the presence of *oh*, the utterance is recognized as a quote, that is, as the reported speaker's utterance not the reporting speaker's; in that function, *oh*'s directionality is forward-pointing. Similarly, *well* in *well I can't feel any* is dual in

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

function: w*ell* as marking a dispreferred response is bi-directional pointing both backward (to the utterance which the response is a response to) and forward (to the dispreferred encoded in the utterance it introduces), while *well* as marking the beginning of a quote is only forward-pointing. So, the two marking functions need not be seen as mutually exclusive: they can be effective simultaneously.[18]

What can be said in support of the interpretation of quote-first *oh* and *well* as quotation markers? The notions of lexical priming and, more specifically, textual colligation (Hoey 2005) can be cited fruitfully. The theory of lexical priming holds that "[e]very word is primed for use in discourse as a result of the cumulative effects of an individual's encounters with the word" (Hoey 2005: 13). There are many conditions that specify a word's 'use in discourse,' such as, for example, its collocations, colligations, and pragmatic associations, and which determine how the word is primed in a user's mind. Another such condition is the word's textual position: "[e]very word is primed to occur in, or avoid, certain positions within the discourse; these are its textual colligations" (Hoey 2005: 13). The textual colligation claim can be applied to interjections in the present context. It is by no means uncommon that depending on its context and co-text of use a single word has two or even more primings; see, for example, Hoey's (2005: 88ff.) discussion of key primings of the polysemous noun 'reason.' Therefore, based on the evidence presented above, it appears defensible to assume two position-related primings for the interjections *oh* and *well*. The first priming is for them to occur in first position in conversational utterances as items by which speakers launch their own utterance. The second priming is for them to occur in first position in quotes as items by which narrators launch the quotation of a reported speaker's utterance. Bearing in mind the above-noted significant overuse of *oh* and *well* in quote-first position (compared to their occurrence in utterance-first position), the priming for quote-launch may be seen as the stronger of the two primings. The assumption of the two-fold priming, however, raises the question of how the two primings can be distinguished by the listener, that is, how the listener can ascertain which of the two primings to activate. The distinguishability seems to be determined by three 'conditions of use.' First, Hoey (2005) repeatedly emphasizes that primings are subject to variation, inter alia, by genre. Priming theory takes account of "who is speaking or writing, what is spoken or written about and what genre is being participated in" (Hoey 2005: 13). The two primings for *oh* and *well* each are genre-specific: the first priming, for utterance-first position, is a priming for general (non-narrative) conversation. The second priming, for quote-first position, is a priming for conversational narrative. In storytelling, where, by definition, quotation is much more likely to occur than in general conversation (see, for example, Mayes 1990, Baynham 1996), the occurrence of Free Direct and Direct discourse presentation and thus the occurrence of quote-first *oh* and *well* are highly expectable. In storytelling,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

the interpretation of *oh* and *well* as signifying the onset of quotation is hence especially relevant. Second, the primings are likely to be distinguishable on a lexical basis. As regards the above cited noun 'reason,' Hoey demonstrates that its multiple primings can be distinguished phraseologically: the rarer senses of 'reason' (viz. 'logic' and 'rationality') "avoid the same collocations, colligations, and semantic and pragmatic associations" (Hoey 2005: 88) associated with the common sense ('cause'). As regards *oh* and *well*, it is likely that similar phraseological distinctions apply, that is, that *oh* and *well* in quote-first position enter into different lexical associations than utterance-first *oh* and *well*. Obvious such associations avoided by utterance-first *oh* and *well* but supported by quote-first *oh* and *well* are, in the case of Direct quotes, the collocation with quotatives (*says*, *said*, etc.) and, accordingly, the colligation with reporting clauses (*he said*, *she says*, etc.). Indeed, these associations are extremely close: of the 168 quote-first instances of *oh*, as many as 142 instances, or 85%, are introduced by a reporting clause, while the corresponding number for *well* (131 instances in quote-first position) is 127 instances, or 97%. For illustration, see almost any of the examples in this section.[19] Third, it can be shown that a straightforward non-phraseological distinction is possible too, viz. via position. For although the two primings have in common the association with first position, they are, at the same time, distinguished by position. This is because quotes occur *within* utterances – namely narrative utterances. Their occurrence *within* the utterance entails that quote-first *oh* and *well* normally do not occur in first position in that utterance but in utterance-*medial* position. (Only in the rare case that *oh* and *well* occur at the immediate onset of a Free Direct quote that *initiates* the narrative utterance do they come to occur both in quote- and utterance-first position; see below.) Excerpt (4.8) illustrates the issue:

(4.8) Yeah alright I just don't know why he doesn't. [^{MSS}He didn't give me satisfactory answer] he said ^{QSD} [^{MDD} **oh** I'll think of it]
(KCV-N1)

As noted, the position of an item is calculated as the proportion of the number of words preceding that item divided by the total number of words in the unit under inspection (utterance or quote). Thus, in (4.8), the position of *oh* within the quote is 0/6 = 0. However, *oh*'s position within the narrative utterance is 21/27 = 0.77.

That quote-first *oh* and *well* indeed occur in *medial* position in narrative utterances, can be demonstrated comparing two data subsets: (i) positions of quote-first *oh* and *well* computed, not relative to the quote (where their position is 0), but to the narrative utterance which encloses them and, as shown above, (ii) positions of *oh* and *well* occurring in first position in conversational utterances. To keep matters simple, only narrative utterances containing exactly one quote were investigated. The distributions of *oh* and *well* across the two sets are depicted in the boxplots in Figure 4.6:[20]

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

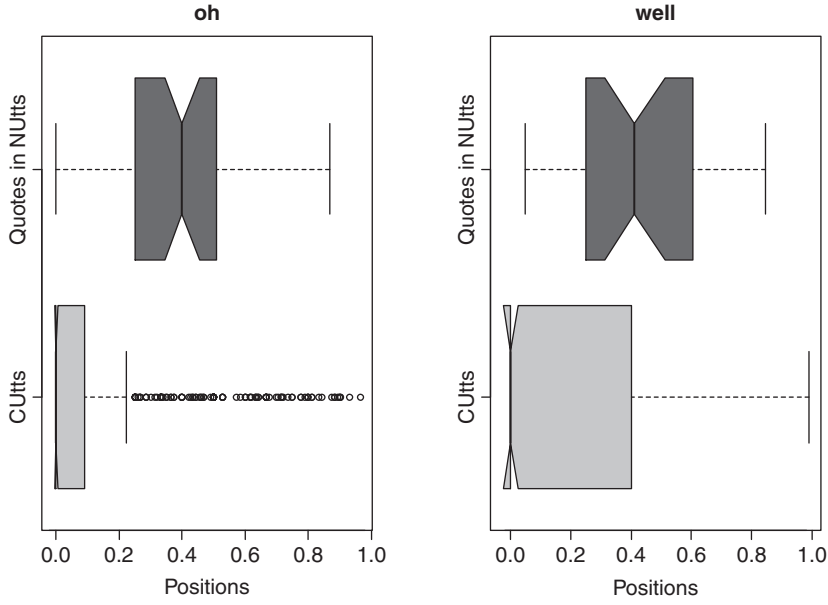


Figure 4.6 Boxplots of positions of *oh* and *well* occurring (i) within narrative utterances (NUtts) in quote-first position and (ii) within conversational utterances (CUtts)

In Figure 4.6, the dark grey boxes represent positions of *oh* and *well* in quote-first position within their enclosing narrative utterances (NUtts), while the light grey boxes represent their (already familiar) positions across conversational utterances (CUtts) (see Figure 4.4 above). First, note that the whiskers for quote-first *oh* (but not for quote-first *well*) in NUtts extend to the value 0. This indicates that indeed *some* instances of *oh* (but not of *well*) occur in first position *both* in the quote and in the enclosing NUtt (i.e., utterance-initially). But they are a minority. The clear majority of instances of quote-first *oh*, and also *well*, occur in positions far higher than 0. This can be seen instantly from the fact that the (dark grey) boxes for quote-first *oh* and *well* in NUtts are located much farther on the right than the (light grey) boxes for *oh* and *well* in CUtts. This visual distance shows up mathematically as the difference between the medians (indicated in the bold vertical lines in the boxes): the median for *oh* and *well* in CUtts, which, as noted earlier, is 0 for either item, contrasts starkly with the median for *oh* and *well* used in NUtts, which is 0.4 (again for either item). The difference is very highly significant: according to a Wilcoxon rank sum test, for *oh*, p-value < 2.2e-16, and for *well*, p-value = 1.147e-07. This is clear evidence that the two primings for *oh* and *well*, viz. for utterance-launch and quote-launch, can be kept apart on positional grounds. Altogether, then, recipients can draw on three factors

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

to determine which of the primings is contextually relevant: genre (quote-launch in conversational narrative vs. utterance-launch in general conversation), lexical associations (e.g., in Direct quotes, quote-first *oh* and *well* collocate with quotatives and colligate with reporting clauses, while these associations are avoided by utterance-first *oh* and *well*), and, finally, position: within the narrative utterance that encloses them, quote-first *oh* and *well* occur, not in utterance-first, but utterance-*medial* position.

Given their distinct, and strong, priming for quote-first position, *oh* and *well* can be seen as signals indicating the left-hand boundary of quotation: narrators use *oh* and *well* to signpost that constructed dialog is immediately forthcoming. The beneficiary of this signaling is the recipient: for them, the inception of quotation is made prominent such that the narrator's 'voice' can be told apart from the characters' voices and the characters' voices can be told apart from one another. In using quote-first *oh* and *well* as quotation markers, narrators design quotation with a view to the recipient's processibility needs. These needs are pressing given the vascillation of reference inherent in the use of (strings of) constructed dialog, and aids to ease processing are therefore particularly welcome.

4.2.5 Summary

This case study on positioning of interjections in quotes as compared to conversational utterances investigated the possibility that interjections serve a quotation-marking function. It was shown that two interjections – *oh* and *well* – occur significantly more often in very first position in quotes than they do in conversational utterances. Two possible interpretations were discussed: that utterances expressing surprise (*oh*) and deviation from conversational expectation (*well*) are preferably pre-selected for quotation and, alternatively, that *oh* and *well* are *primed* for, not only utterance-first (in general conversation), but, even more strongly, quote-first position (in narrative), thus serving an additional discourse-deictic role of flagging the left-hand boundary of presented discourse. The two interpretations are by no means mutually exclusive: why should presenters not both pre-select certain utterance types that promise to be of interest to the recipient and, at the same time, value the discourse-structuring effect of certain interjections that typify these utterance types? Also, the two interpretations align to a larger principle, the principle of recipient design: no matter whether *oh* and *well* are interpreted in the context of pre-selection or as left-boundary quotation marks, their use in quotes will primarily serve the recipient: their interest (in drama) and need (for processibility). Seen thus, as means in the service of recipient design, the statistical overuse of the small items *oh* and *well* in quote-first position contributes to the large enterprise binding narrator and recipient together: that of co-constructing narrative in conversation.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

While this section has investigated positioning in the use of verbal items, viz. interjections, the next section targets the positioning of non-verbal items, viz. unfilled pauses.

4.3 Pauses as quotation markers

4.3.1 Introduction

As shown in the previous case study, certain interjections in first position play a central role in marking quotation thus aiding the recipient in separating the narrator's own words from some reported speaker's words. However, interjections do not occur in all quotes and even where they do, they are not always used in first positions. So very likely there will be other resources at work too that speakers draw on to flag discourse as quoted. Some such markers have been identified in previous research. Besides gestural markers such as 'air quotes' and raising eyebrows (Brendel *et al.* 2011: 9), mostly paralinguistic marking resources have been taken notice of. They include changes in voice quality or prosody (Klewitz & Couper-Kuhlen 1999, Mathis & Yule 1994: 76, Holt 1996, 2007: 57), particularly for Free Direct quotes, and, at least for Russian conversation, pauses (see Bolden 2004). To my knowledge, pauses and, specifically, silent pauses have not yet been investigated in previous research with a view to their potential as quotation markers in English conversational narrative. The aim in this section is to provide this as yet missing investigation.

While pauses, particularly in Chomskyan linguistics, used to be seen in a negative light as performance errors (see for example, Chomsky 1965: 49) and were often regarded "as a nuisance, as a kind of debris lying in the way of ordered exposition" (Kjellmer 2003:170), more recent research, particularly in corpus linguistics, has discovered vital roles for pauses in discourse and interaction; for pauses in turntaking, see, for example, Stenström (1990), Biber *et al.* (1999), Wennerstrom (2001), Kjellmer (2003), and Rühlemann (2007); for the relation of pauses with information structuring and speech planning, see, for example, Chafe (1992: 91), Allwood *et al.* (1990), Stenström & Svartvik (1994), and Biber *et al.* (1999: 1054). Generally, there seems to be some agreement that pauses occur at *important junctures* in the discourse. In Biber *et al.*, for example, silent pauses tended to "occur at major boundaries between syntactic units" (Biber *et al.* 1999: 1054), while Chafe (1992: 91) noted a co-occurrence of pauses with tone unit boundaries, which he sees as prosodic demarcators for new ideas.

The specific roles of pauses in narrative have been much less researched. Chafe (1987) found for pauses an information-structuring function in narrative: in his analysis (of a single story, though) pauses were interpreted as markers of what he called 'spoken paragraphs' including 'summary,' 'instantiation,' and 'wrap-up.' Rühlemann *et al.* (2011) is, to my knowledge, the first

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

corpus-based investigation of pauses in conversational narrative. The study reports a close association of pauses and quotative verbs used to introduce Direct discourse presentation. It was shown that the filled pauses *er* and *erm* as well as short silent pauses (up to five seconds) occur significantly more often in initial positions in Direct quotes than in Indirect reports.

The following case study extends and deepens this analysis. Two methodological differences, however, need to be noted. First, Rühlemann *et al*. investigated short silent pauses as well as the filled pauses *er* and *erm*. Given that short silent pauses by far outnumber filled pauses in quotes[21] (see below), I will focus on short silent pauses and refer to filled pauses only for comparative purposes. Second, due to their interest in the association of pauses with *quotatives*, the authors did not consider pauses in Free Direct quotes (which are, by definition, without quotatives). In the present research, by contrast, I will consider short silent pauses both in Direct and in Free Direct quotes.[22]

Another word is necessary regarding the type of positioning investigated. Unlike the previous case study, which examined interjections in first position, the present study will explore the positioning of pauses *across quotes overall*, for two reasons. First, while interjections are verbal material, the nature of silent pauses is non-verbal and paralinguistic. Given its paralinguistic nature, there is in fact little reason to assume that the 'job' of marking quotation must be accomplished by pauses precisely at the boundary between the presenter's own discourse and the presented discourse. As paralinguistic markers, it seems, they may more easily skip discourse boundaries than, for example, lexical items. In the context of marking quotation, it is therefore hard to see why narrators should not also attend to that marking task by using pauses in the boundary's *vicinity*, where they could still accomplish their purpose of flagging the narrator's words as somebody else's. Secondly, pauses are distinguished from interjections also by their differential behavior vis-à-vis positioning. We saw above that interjections are very heavily attracted to utterance- and quote-beginnings. No such positional preference has yet been noted for pauses. Based on the above cited agreement that pauses occur at important junctures in the discourse and given that the inception and completion of a quote constitute prime examples precisely of such junctures, it seems consistent to assume that pauses will occur not only quote-initially, like interjections, but also quote-finally. If pauses were indeed found to *frame* quotation in this way they could be seen as alleviating not only the left-hand boundary issue, like interjections, but also the right-hand boundary issue; that is, pauses would also attend to the problem of flagging the narrator's return from the reported speaker's words to their own words or their re-orientation to a second reported speaker's words.

I will thus test the following hypotheses:

H_o: Short silent pauses do not occur more frequently in initial and final positions than in medial positions in quotes.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

H_1: Short silent pauses occur more frequently in initial and final positions than in medial positions in quotes.

How are initial and final positions defined? I define as initial positions those positional values that range from 0 to 0.2, mathematically: $\epsilon[0,20]$, and as final positions those values ranging from 0.8 to 1, mathematically $\epsilon[80,100]$.

4.3.2 Data and methods

Given the non-verbal nature of silent pauses, a central question is whether they can be considered an integral part of the quote. The issue at stake is illustrated in (4.9) for pauses occurring at the left-hand boundary of quotation; the issue is of course the same for pauses occurring at the right-hand boundary (see (4.10) below) and occurring in quote-medial position:

(4.9) I thought **<pause/>** you shouldn't speak to people like that (KB0-N2)

The silent pause in (4.9) sits precisely between the reporting clause *I thought* and the reported clause (the quote) *you shouldn't speak to people like that* (that is, between two 'major syntactic units'; see Biber *et al.*'s (1999: 1054) relevant observation). Which part is it best attributed to: the narrator's discourse or the quoted discourse?

Quotation is by no means restricted to *speech* as the only mode of communication to be quoted. Research on BE *like* and GO, for example, has amply demonstrated the quotatives' ability to introduce a broad range of different types of 'discourses', including not only speech and thought (like SAY and, respectively, THINK), but also gesture (Butters 1982: 305, Ferrara & Bell 1995: 281), and emotion (Romaine & Lange 1991: 238, Ferrara & Bell 1995: 282ff., Buchstaller 2002: 15, Rühlemann 2007: 149ff.). So, it cannot be ruled out that presenters expressly quote pausing. Nonetheless, it seems counter-intuitive to assume that the majority of pauses occurring inside quotes or at their boundaries are *objects* (and hence part) of what the presenter is intending to present. Rather, it seems, they are best seen as expressions of what Allwood *et al.* (1990) refer to as 'speech management,' that is, "externally noticeable processes whereby the speaker manages his or her linguistic contributions to the interaction and to the interactively focused informational content" (Allwood *et al.* 1990: 3). Seen as speech management expressions, pauses would have to be placed on a discourse plane outside the quote. That is, to return to (4.9), by use of the pause the narrator separates the reporting clause *I thought* from the quote; the pause itself is not part of the thought presentation. It marks the boundary between the two units.

How are pauses treated in the present analysis of their positioning in the context of quotation? Considering the aim of the present study – to discover whether pauses frame quotation – the two boundary types of pauses were

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

included within the scope of the quote although, in fact, they may have been placed outside the <seg>-element delimiting Direct and Free Direct quotes. Re-consider (4.9), printed here as (4.10), which gives a little more context as well as the annotation for reporting mode:

(4.10) "Bloody ignorant pig" (Type: T10/embedLevel: EC1)
(. . .)
S1 PNP I thought **<pause/>** [^{MDD} you shouldn't speak to people like that] but **<pause/>** [^{MRR} I'd said it]
(KB0-N2)

It can be seen that the pause between *thought* and *you* is placed outside the quote's boundaries defined by the annotation. However, note also that no verbal material (that is, in annotational terms, no <w>-element) intervenes between the pause and the quote. Therefore, for the purposes of the present research, pauses like the first in (4.10) were *included*, not within the quote (as said earlier, pauses cannot be seen as objects of quotation), but within its *positional* scope. That is to say that pauses occurring *immediately before* the quote were assigned the positional value 0, that is, the same value that would be obtained if the pause had been included inside the quote's annotation. By contrast, pauses occurring before the quote but separated from it by intervening verbal material were excluded. By the same token, pauses occurring *immediately after* the completion of a quote were included within the quote's positional scope, whereas after-quote pauses distanced from the quote by inbetween word elements were excluded. For example, in (4.10), the pause following the quote was discounted given the intervening conjunction *but*, which signals the inception of a new clausal unit. Note that for silent pauses occurring at the extreme right end of the positional scope (that is, as quote-last element or as the very first element after the quote), the positional value is 1. This is because, as noted in Section 4.2, underlying the calculation of positions is the division of the number of words preceding the item whose position is to be determined by the total number of words in the unit (the quote, in this case). Silent pauses, however, do not represent words but XML elements of their own, which, in XPath terms, are treated as siblings to word elements (<w>) (see Section 2.3.2). They are hence not counted as words. Consider (4.11): the quote contains 18 words. The position of the pause following the quote is 18/18 = 1, a value which adequately reflects the item's quote-boundary position.

(4.11) "Microwave" (Type: T10/embedLevel: EC2)
(. . .)
S1 PNP (. . .) And she ^{QSD}said [^{MDD} no you have it my love. I've just bought a new one and you can have it] **<pause/>** (. . .)
(KB6-N1)

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

Table 4.6 *Frequencies of* er, erm *and* <*pause*/> *occurring within the positional scope of quotes*

	Freq	%
<pause/>	362	77
er	63	13
erm	47	10
Total	472	100

To keep the terminology simple, I will refer to positions in the positional segments ϵ[0,10] and ϵ[90,100] as 'quote-initial' and 'quote-final' positions although pauses occurring immediately before and, respectively, after quotes do not strictly belong to the quoted material. As is usual in this book, XQueries were used to calculate positions of pauses.[23] Finally, the transcription in the BNC-C of paralinguistic features such as silent pauses may not be perfectly consistent (see Hoffmann & Evert 2007, Hoffmann *et al.* 2008: 57).[24] Therefore, the following results should be treated with some caution.

4.3.3 Results

At first, it seems appropriate to establish how frequently silent pauses occur within the above-described positional scope of quotes. For comparison, Table 4.6 also presents the observed frequencies of the filled pauses *er* and *erm*, by far the most frequent forms (other potential forms such as *ehm* or *uhm* are not attested in the NC).

Short silent pauses are by far the most frequent pause types in the data: the 362 instances of <pause/> account for far more than two-thirds of all pauses (77%) , while the 63 *er*s and 47 *erm*s combined take up the remaining 23%. So pausing in the context of quotation is indeed mainly a non-verbal phenomenon and we are justified in restricting the focus of this research to silent pauses. However, considering that the total number of quotes in the NC is 1,784 (see Section 2.2.2), the 362 silent pauses in the data are involved in less than a fifth of quotes overall ('less', because, as will be shown below, many quotes have two pauses, one on either end).

How do pauses distribute positionally? If the above-noted assumption is correct (that silent pauses are more frequent in initial and final positions than in medial positions) we will expect to find a U-shaped distribution across the quotes' positional scope as defined above; that is, a distribution will result that has exactly two peaks: one for initial positions, defined as the positional segment ϵ[0,20], and one for final positions, defined as ϵ[80,100].

Figure 4.7 shows the distribution of positions of silent pauses and of the filled pauses *er*/*erm* as histograms. It will be immediately obvious from the

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

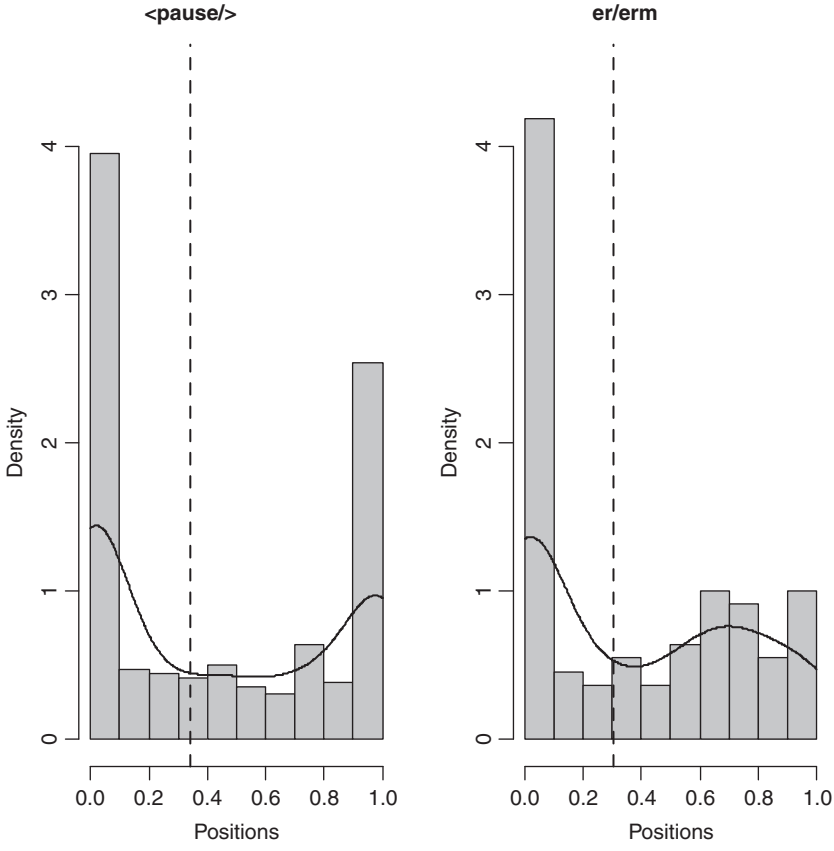


Figure 4.7 Histograms of positions of silent pauses (<pause/>) and the filled pauses *er* and *erm* taken together within the positional scope of quotes

graph in the panel on the left that silent pauses behave almost exactly in the way predicted, namely with the highest frequencies of occurrence in initial and final positions, resulting in a density curve which indeed resembles a U and peaks located exactly in the two critical positional segments ϵ[0,10] and ϵ[90,100]. However, the peaks are of unequal size: the left-hand peak is greater than the right-hand peak. This distribution suggests that the probability for silent pauses to occur quote-medially is smaller than to occur quote-initially and also, to a lesser extent, smaller than to occur quote-finally. By contrast, a U-shaped type of distribution is not found for *er/erm*:[25] while the filled pauses too have their primary peak in initial positions, their secondary peak is not, as expected, in final but pre-final positions, more specifically in the range ϵ[60,80]. Contrary to expectations, in the final

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

segment the curve even reaches, not its second high point, but its lowest point. This distribution suggests that *er/erm* are less likely to occur quote-finally than pre-finally and much less likely than initially.

To what extent are the above observations significant? Since the focus is on silent pauses, the question of significance will only be addressed for this type of pause. That is, to what extent can significance be claimed for the preferences for initial and final positions observed for short silent pauses? The 'natural' answer to this question would be to perform an appropriate statistical test. However, significance tests essentially involve juxtaposing the distribution of a given variable against the distribution of another comparable variable or against a specific distribution (e.g., normal distribution, exponential, etc.) or against an a priori stipulated value (Dalgaard 2008: 95). In the present case, however, we have one sample only and no a priori benchmark value against which to compare the positioning of silent pauses. One way to nonetheless make reliable inferences is the non-parametric density estimation and the use of the bootstrap to compute confidence intervals. For reasons of space, this method cannot be discussed here in great detail (for a readable introduction, see Mooney & Duval 1993); some basic characterization must suffice.

The bootstrap is a non-parametric resampling method introduced in Efron (1979). In linguistics, it is used occasionally in computational linguistics (e.g., Erk *et al.* 2010) and machine translation (e.g., Koehn 2004) and is beginning to be used in corpus research (e.g., Gries 2006, Lijffijt *et al.* 2011). Underlying it is the "approach to treat the sample as if it is the population" (Mooney & Duval 1993: 9). From the sample of size n, a large number of resamples are drawn randomly. Crucially, the resamples are of the same size n as the original sample and are drawn *with replacement*. Therefore, each resample may contain some sample data points more than once whereas other sample data points are not included at all. As a result, "each of these resamples will likely be slightly and randomly different from the original sample" (Mooney & Duval 1993: 10). These differences have repercussions: the mean and the standard deviation of each resample vary from those of one another and from those of the original sample.[26]

Why generate resamples? Generating a large number of resamples with slightly different kernel density estimations enables the calculation of pointwise standard errors for the estimated density and, hence, the calculation of confidence intervals (CIs). These are crucial to the method in that the CIs facilitate statistical inference (see Section 3.2). Now, crucially in the present context, CIs can be visualized by confidence bands. Confidence bands are obtained from a six-step procedure (executed in R): (i) draw a large number of resamples, (ii) for each resample estimate the density in 512 points (a number predefined in R) dispersed across the minimum and the maximum values in the data (that is, in the case of positional values for pauses, between 0 and 1), (iii) calculate the mean for each of the 512 points over the resamples,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

(iv) calculate from all resamples for each point one single standard error (a measure of how widely the density estimations vary), (v) calculate on the basis of the standard errors the 512 confidence intervals, and (vi) connect them such that they form the confidence band: its bounds delimit all those density estimates that cover in 95% of all cases the true density estimate of the population in question.

The confidence band obtained from the procedure invites interpretation. Crucial to its interpretation is its width. It determines the precision and hence reliability of the density estimate for the original sample: the narrower the band, the greater the precision and reliability; the broader the band, the smaller the precision and reliability. Since the bounds of the band delimit all those density estimates that cover the true density estimate in 95% of the cases, any conceivable curve representing any conceivable estimate within the band could, in principle, represent the true estimate. If the upper and the lower bounds of the confidence band are so far apart that between them curves are conceivable whose shapes are essentially different from the density curve obtained for the sample, then the shape of the density curve for the sample is an unstable and unreliable foundation for any inference as to the shape of the curve to be obtained from the true population. Such an essentially different shape would obtain, for example, if any two points within the confidence band could be connected in such a way that a completely new primary maximum would arise at a point along the x-axis where, in the original sample's density curve, there is no such maximum. Such a scenario is the more likely the wider the confidence band. If, by contrast, the width is narrow such that alternative contours would merely modify the existing contour in negligible details, then the ground on which to infer from sample to population is stable and reliable: we can be 95% confident that the microcosmos of the finite sample at hand is a faithful mirror image of the macrocosmos of the population.

For a reliable bootstrap, a sample size of 30 values is often considered a sufficient basis (see Mooney & Duval 1993). The sample size of 362 silent pauses and their positional values is thus more than sufficient. On this sample, a bootstrap was performed with 5,000 resamples drawn randomly from the original sample; the significance level α was set to 0.05. The results can be inspected in Figure 4.8. The figure shows the same histogram and the same density curve for positions of pauses in quotes as in the left-hand panel in Figure 4.7 above. The dashed lines below and above the density curve delimit the confidence band calculated on the basis of the bootstrap.

As said above, the interpretability of the confidence band depends on its width. If the band is very wide, admitting shapes of density estimates largely different than the density estimate for the original sample, then inferences made on the basis of the sample's curve are hardly reliable. If, by contrast, the band is narrow admitting at best variations of the existing curve, then the curve is meaningful in terms of statistical inference. In the case of the

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

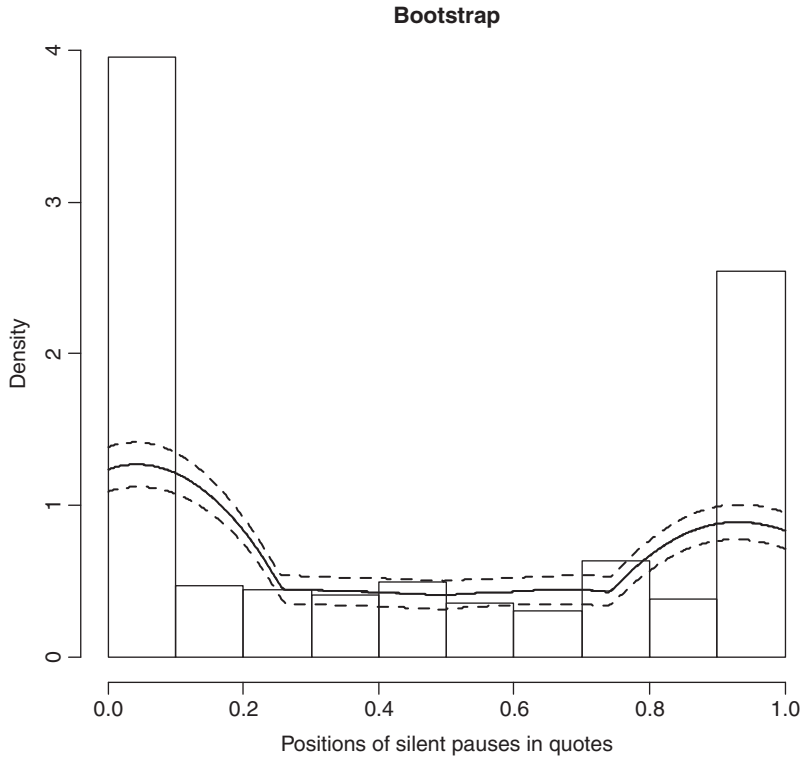


Figure 4.8 Histogram of positions of silent pauses in quotes with confidence bands calculated on the basis of a bootstrap (B = 2000, $\alpha = 0.05$)

confidence band shown in Figure 4.8, the latter scenario is clearly the case: the band is markedly narrow, particularly around the values 0.25 and 0.75. At these points on the x-axis, any alternative curve must follow, on the left, the original curve's decline from its primary maximum between 0.0 and 0.1 as well as replicate, on the right, the curve's rise towards its secondary maximum between 0.9 and 1. Around the peaks, the band is a little wider. This indicates that, here, alternative contours are conceivable which peak not, as in the original sample, between, on the left, 0 and 0.1 and, on the right, between 0.9 and 1, but between 0.1 and 0.2 and, respectively, between 0.8 and 0.9. However, peaks beyond 0.2 and, respectively, 0.8 are not conceivable. Further, the highest point of the valley between the peaks is still much lower than the lowest point of the primary maximum on the left and the lowest point of the secondary maximum on the right. Even if the true density estimate covered these extreme points, the overall shape of the curve would form a U: a (slightly lower) primary maximum for initial positions and a

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

(slightly lower) secondary maximum for final positions. The original sample's maxima would still be firmly in place and the overall contours of any alternative density curves would still be the same as the contour of the density curve obtained for the sample. That is, the confidence band calculated on the basis of the bootstrap suggests that the true density estimate for short silent pauses in English narrative will essentially resemble the density estimate calculated for the sample drawn from the NC, exhibiting peaks for initial positions and final positions and a valley for middle positions. This is conclusive evidence, then, to suggest that the alternative hypothesis according to which short silent pauses are more frequent in the positional segments $\epsilon[0,20]$ and $\epsilon[80,100]$ than in any positional segments in between, can be accepted.

4.3.4 Discussion

How to interpret these positional preferences for initial and final positions of short silent pauses? Two possible interpretations come to mind. Rühlemann *et al.* (2011) propose a speaker-centered interpretation that links the use of pauses in quotation to processing constraints imposed on the presenter by the necessity in quotation to perform a "*shift* into a non-present speaker's voice and reference system" (Rühlemann *et al.* 2011: 225). This interpretation is in line with Allwood *et al.*'s (1990) view of pauses as choice-related speech management expressions, serving "to enable the speaker to gain time for processes having to do with the continuing choice of content and structured expression" (Allwood *et al.* 1990: 10–11). Rühlemann *et al.* (2011) were concerned only with initial positioning of pauses. They were not aware of the fact that silent pauses also peak at final positions in quotes. This oversight has reverberations for the interpretation of how pauses position. If it is assumed that pauses peak only at the onset of quotation the speaker-centered interpretation sounds highly plausible: the shift from the speaker's own default reference system into a 'persona's' reference system is likely to incur high costs in terms of processing. Rühlemann *et al.* also noted the costs of assuming a character's '*voice*' system. If voice is understood in a narrow sense as referring to mimicking intonation and voice quality, the change from the speaker's own voice to a character's voice will incur further processing costs. The double switch from one's own reference and voice system to a persona's reference and voice system will require some planning time; hence the use of silent pauses. However, the present research has discovered that silent pauses also preferably occur upon the completion of quotes – that is, at points in the narrator's discourse where, under most circumstances, a return can be observed to the narrator's own words and thus his/her own reference and voice system (only where the quote is directly followed by another quote is there no such return to the narrator's own words but instead a switch into yet another reported speaker's reference and voice system). It seems less

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

plausible to assume that the return to one's own system, which can be considered the default system, is as processing-intensive as the switch into an alien system. Thus, the speaker-centered interpretation of pauses as planning buffers is likely not sufficient alone.[27]

A complementary interpretation can be arrived at if the use of pauses in the context of quotation is approached from the perspective of the recipient. Given the non-verbal nature of silent pauses, their occurrence within discourse causes a temporary absence of discourse which "interrupts the articulatory flow of speech" (Klewitz & Couper-Kuhlen 1999: 476). Given their preferred recurrence at both ends of quotation, pauses effectively divide discourse into pre-pause and post-pause discourse thus 'cutting' it into segments: the quote as a whole is exposed from the flow of discourse as a discourse segment with discernible boundaries. The segmentation alerts the recipient to the possibility that the narrator is preparing or concluding a major transition of some kind: either a transition from one syntactic unit to another (see Biber *et al.* 1999), or a shift from 'one new idea' to another new idea (see Chafe 1992) or, as in the case of quotation, shifts in the narrator's perspective: (i) from author formulating their own discourse to animator creating a likeness of someone else's discourse, or (ii), the reverse, from animator to author, or (iii) from animator of one character to animator of another character, as is the case in quotes occurring in strings (see Section 5.2). Seen in this recipient-centered perspective, pauses inevitably serve as quotation markers. Evidence to support this interpretation is the U-shape of the distribution: silent pauses *bracket* quotes, most typically occurring right before and right after the quote, thereby strikingly resembling quotation marks used in writing. If pauses mainly or only served to buy the narrator planning time we should expect that pauses tend to occur quote-initially only. Their recurrence quote-finally is much less expected given that the narrator's return to their own words, reference, and voice system is likely to require much less planning and hence much less pausing. That is, the tendency for silent pauses to bracket quotation fulfills a strategic function in the service of recipient design: the quote, as a whole, is separated from the neighboring discourse, marking its onset and completion and alleviating the boundary issue at either end. Pauses are thus designed as a quote-unquote signal – a design supporting the recipient's processibility needs vis-à-vis the narrator's multiple changes in footing.

The two views complement each other, in two respects. High-frequency items generally and non-verbal items specifically are prone to perform multiple functions. Therefore it seems not unlikely that narrators use pauses both as planning buffers for themselves *and* as quotation markers for the recipient. Even if narrators did not make use of pauses strategically, that is, as a means of reaching the goal of aiding the recipient in processing quotation (see van Dijk & Kintsch 1983: 65), the use of pauses, as a temporary absence of discourse, would still have the convenient side effect for the recipient to

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

divide the narrator's discourse into discourse chunks. Given that the chunks' boundaries are far from random but coincide with discourse boundaries – viz. the onset and completion of quotation – the chunks as such are meaningful, thus facilitating the identification of quotation and, as a consequence, aiding its processing.

4.3.5 Summary

In this section I have investigated the role of short silent pauses in the context of quotes. The data considered included both pauses occurring within the annotational boundaries of quotes as well as pauses occurring immediately before or after the quote. Starting from observations in the literature that silent pauses preferably occur at major syntactic and discourse boundaries it was hypothesized that the distribution of pauses forms a U, with high rates of occurrence in initial and final positions. Initial positions were defined as positional values in the segment $\epsilon[0,20]$, while final positions were defined as positions in the range $\epsilon[80,100]$. The positional analysis demonstrated that indeed silent pauses preferably occur at the onset and, respectively, upon completion of the quote. It was noted in passing that no such U-shaped distribution could be found for the filled pauses *er* and *erm*, which exhibited a peak in initial positions but no corresponding peak in final positions. To establish the significance of the U-shaped distribution of silent pauses, a bootstrap was performed calculating a confidence band for the distribution. The confidence band thus obtained turned out to be extremely narrow leaving no room for any essentially different density estimate than the density estimate obtained from the original sample. The alternative hypothesis was therefore fully accepted: silent pauses frame quotes, with peaks in initial and final segments.

Two alternative interpretations were discussed. In a speaker-centered perspective, pauses were seen as speech management expressions serving to buy the narrator planning time, a function which is vital given the considerable processing costs incurred by the fact that quotation requires a shift from the narrator's own reference and voice system to a character's reference and voice system or from one character's to another character's systems. It was noted that the fact that pauses *frame* quotes, peaking not only at the beginning but also at the end of quotation, is less well explained by the view of pauses as indices of the narrator's planning constraints in that the return to the narrator's own reference and voice system arguably requires less such planning. In a recipient-centered perspective, it was argued that silent pauses, as temporary absences of discourse, have the effect of separating pre-pause discourse from post-pause discourse. Given their separating effect, pauses can be seen as a convenience to the listener, aiding their understanding that, in the case of quote-initial pauses, the narrator is about to embark on quotation and, in the case of quote-final pauses, the quote has been

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005

completed. Seen in this perspective, the use of silent pauses by narrators follows the macro-strategy of recipient design: pauses dissociate quotes from the 'stream of narration' thus flagging both the inception and completion of quotes. As markers of quotation, pauses are valuable processing instructions for the recipient aiding them in resolving the oscillation of reference that is the hallmark of quotation, particularly where quotes occur in strings.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 21:59:51, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139026987.005