

Monsters Aren't Enough: (Non-)Indexical Shift in Indirect Discourse

Sebastian Walter, Goethe University Frankfurt
Stefan Hinterwimmer, University of Hamburg

Since Kaplan's (1989) prohibition of so-called monsters, it has been widely assumed that indexical expressions, such as "I", "here", or "tomorrow", can receive shifted interpretations, i.e., non-speaker-oriented interpretations, only in direct discourse, but not in indirect discourse. However, subsequent work has challenged this view by showing certain perspective-dependent expressions can take their reference from the reported rather than the reporting context (Schlenker 2003; Anand & Nevins 2004; Deal 2017; Saure 2025), which has led to a surge in monstrous analyses of indexical shift. Building on this line of research, the present paper extends the empirical domain of perspective shift in indirect discourse to multimodal meaning. The results of a rating study reported in Walter & Hinterwimmer (2025b) show that face emoji in indirect discourse can be interpreted from the perspective of the matrix subject rather than that of the author. We analyze this finding in terms of mixed quotation. We then extend this non-monstrous analysis to cases of shifted indexical adverbials in indirect discourse, thereby limiting our focus on languages that are normally argued to disallow indexical shift. We argue that this analysis is more suitable than a monstrous analysis of these findings because indexical shift in these languages must be licensed pragmatically. Therefore, an operator-based, monstrous analysis would be too strong.

Keywords: *indirect discourse, perspective-taking, indexical shift, emoji, multimodality*

1. Introduction

Recent years have seen a growing interest in how perspective-taking operates across different modalities of language, including not only speech but also writing and visual means such as emoji, gesture, as well as the interactions of several modalities with regard to perspective-taking (e.g., Walter & Hinterwimmer 2025a). In digital discourse, face emoji have become a key device for expressing affective meaning. Previous studies (Grosz et al. 2021, 2023) have argued that face emoji are by default author-oriented and thus resistant to perspective shift, predicting that they should not be able to receive non-author-oriented interpretations in embedded contexts such as indirect discourse. Despite these claims of interpretive rigidity, related work on perspective shift in language more broadly suggests that perspective-sensitive expressions need not always be fixed to the speaker's context. Studies on temporal and local indexicals in indirect discourse (Plank 1986; Anderson 2019; Saure 2025) have shown that even in languages like German or English—which are normally assumed to prohibit shifted interpretations of indexicals outside direct discourse—shifts in interpretation can occur when the reported speaker's perspective is salient.

In this paper, we build on these observations and extend them to the multimodal domain. We argue that, under certain conditions, face emoji in indirect discourse can also be interpreted from the perspective of the matrix subject rather than that of the message author. This suggests that perspective shift in indirect discourse is not confined to verbal expressions. To capture shifted interpretations of face emoji in indirect discourse, we propose that it involves mixed quotation, that is, partial quotation of the reported speech, that is instantiated

as an instance of demonstration in the sense of Clark & Gerrig (1990), where speakers iconically depict aspects of a prior (speech) event. In this way, we treat shifted emoji as demonstrations that reproduce form-related features of the reported speech event. We will then also show that this analysis easily generalizes to instances of shifted indexicals in indirect discourse.

We want to raise awareness for the following, partly interrelated, theoretical and empirical claims in this article: first and foremost, we want to emphasize that indexical shift is not as restricted as often assumed in the literature. Even in languages like German, a language that is normally considered a non-shifting language, cases of shifted local and temporal indexicals have been attested (Plank 1986; Saure 2025). Moreover, thereby extending observations made in Saure (2025), we would like to argue that perspective shift in indirect discourse is possible not only in literary environments, but also outside of them, even in situations that are clearly text messages, a case where Saure (2025) ruled out the possibility of perspective shift. Our most prominent argument from a theoretical point of view is that cases of perspective shift—at least in default non-shifting languages like German—is best analyzed under the assumption that mixed quotation is (optionally) available in indirect discourse.

The remainder of this paper is structured as follows: Chapter 2 will introduce the relevant background on speech and thought representation, thereby focusing on the divide between direct discourse, indirect discourse, and free indirect discourse. In this regard, traditional analyses of each type of speech/thought report will be discussed. We will then zoom in on indirect discourse further, thereby highlighting that while there exist several analyses for direct and free indirect discourse, there is comparably little theoretical work on indirect discourse. Chapter 3 provides some background on quotation. Crucially, we will adopt the view that quotation is an act of demonstrating, meaning an iconic act depicting features of its referent (Clark & Gerrig 1990). In Chapter 4, we will discuss two different formal approaches in emoji semantics: the lexical one proposed in Grosz et al. (2023) and the pictorial view defended in Maier (2023). Chapter 5 first turns to our proposal for shifted face emoji in indirect discourse (5.1). Integrating the idea of quotation as demonstration into our semantic framework (cf. Davidson 2015), we argue that shifted face emoji are essentially quoted facial expressions from the reported speaker. §5.2 extends the analysis to cases of shifted local and temporal indexicals in German indirect discourse utterances. Chapter 6 concludes.

2. Speech and thought representation and its semantic analysis

2.1 *The traditional distinction between DD, ID, and FID*

In research on speech and thought representation¹, three canonical report types are traditionally distinguished: i) *direct discourse* (DD), ii) *indirect discourse* (ID), and iii) *free indirect discourse* (FID). While they all serve the same main purpose, i.e., to report speech or thought, they still differ with regard to many important respects, such as, for example, perspective prominence of the reported speaker and their distribution (FID is normally confined to

¹ Throughout this article, we will often be sloppy and only mention speech representations or reports. If not indicated otherwise, by that we always mean speech and thought representations.

narratives, DD and ID frequently occur outside of narratives, as well). More generally, they also differ in the extent to which the context of the reported utterance interacts with that of the reporting clause. These differences have important consequences for perspective-dependent expressions in general and indexicals, such as first-person pronouns (*I*) and temporal and local deictic expressions (*now*, *today*), in particular. Indexicals are assumed to receive their semantic value directly from features of the utterance situation, i.e., the *context*, (Kaplan 1989, et seq.), such as the speaker, time, and place of utterance, and whose reference is therefore fixed before the semantic content of the remaining expressions of an utterance is determined. Kaplan (1989) furthermore imposes a prohibition of context-shifting operators outside of quotation—operators he calls *monsters*—thus claiming that indexicals must be interpreted in the context in which they are uttered (though see Schlenker 2003; Anand & Nevins 2004; Anand 2006; Pearson 2012; Deal 2017, 2020, among others, for evidence against this claim).

When reporting an utterance in DD, the *reporting* speaker reproduces another speaker's (*reported* speaker, henceforth) earlier utterance from that speaker's point of view. In formal linguistics, this utterance is assumed to be a *verbatim* repetition of what the reported speaker said or thought (e.g., Quine 1940; Davidson 1979). As we will address below in more detail, however, his assumption is not undisputed (see, among others, Wade & Clark 1993; De Brabanter 2017). In written language, a stretch of DD is typically enclosed by quotation marks:

- (1) Yara said, "I will meet you here tomorrow."

Important for the present purposes is the behavior of the indexicals *I*, *you*, *here*, and *tomorrow* in the sentence enclosed by the quotation marks. The first-person pronoun *I* is interpreted as referring to the reported speaker (i.e., Yara) and the second-person pronoun *you* is interpreted as referring to the addressee of her utterance, not the reporting speaker's addressee. Turning to the local and temporal indexicals (*here* and *tomorrow*, respectively) next, they are also clearly interpreted as referring to place where the reported utterance took place as well as the day following the day of the utterance. Thus, all indexicals receive their semantic value from the context of the reported speech event, not from the context of the reporting speech event, which is predicted in a Kaplanian account, as it is normally analyzed as an instance of pure quotation (cf. Davidson 1979; Partee 1973). Consequently, the perspective of the reported speaker is very prominent in stretches of DD.

Another means to report Yara's utterance is given in (2).

- (2) Yara said that she would meet them there on the following day.

The example in (2) is an instance of ID. Here, the reported proposition is embedded under the propositional attitude verb *say*, the standard assumption being that the reporting speaker must only be faithful to the propositional content of the reported speech event, but not its form. ID is therefore not treated as quotational under standard accounts. Indexicals in ID reports are by default anchored to the reporting speaker, which is why all indexical expressions have been replaced by non-indexicals that are normally treated as anaphoric elements. This has led to its standard non-quotational analysis. Instead, following other propositional attitude verbs (cf. Hintikka 1969), the standard analysis of ID treats verbs of saying and thinking as modal quantifiers, that is, quantifiers over possible worlds. An alternative approach, to be discussed in more detail below, analyzes ID within an event-semantics framework (Maier 2018). Note,

however, that numerous languages have been reported that allow for so-called *shifty indexicals*—indexicals that are interpreted from the context of the reported speech event—in ID reports (e.g., Schlenker 2003; Anand & Nevins 2004; Anand 2006; Pearson 2012; Deal 2020), establishing a clear-cut line between languages that allow for indexical shift, such as Zazaki and Slave, and languages that do not allow for indexical shift, such as German and English. Interestingly, however, it has been noted in the literature that some indexical expressions can receive a shifted reading even in non-shifting languages. Although not directly classified as indexical shift, Plank (1986) has noted that local indexicals in German can receive shifted interpretations in ID. In a similar vein, Anderson (2019) presents empirical support that the temporal indexical *tomorrow* in English can receive a shifted interpretation in ID, as well. Thus, the divide between non-shifting as opposed to shifting languages does not seem to be as clear-cut as is often assumed by theorists.

Finally turning to FID, this report type introduces another speaker’s speech or thought (more often a thought) without any overt marking. An example with a context sentence is given in (3).

- (3) Yara was excited to finally re-unite with her best friend from high school after ten years. She would meet them here tomorrow.

The second sentence is the FID utterance. Interestingly, as has often been argued to be a hybrid of DD and ID (Banfield 1982; Schlenker 2004; Sharvit 2008; Maier 2015), as it shares important parallels with the two other report types when it comes to the behavior of indexical expressions. As can be noted from the example in (3), all indexicals except pronouns and tense markings receive shifted interpretations in FID, meaning that *tomorrow*, for example, refers to the day following Yara’s speech or thought event, which strongly patterns with the behavior of indexicals in DD. Conversely, however, the third-person pronoun *she* must be used to refer back to Yara, making it also parallel to ID.

2.2 FID and the idea of hybridity

The “hybrid” conception treats FID as the simultaneous activation of two deictic centers: that of the reported and that of the reporting speaker. Its mixed behavior with respect to the interpretation of indexicals led to analyses invoking double-context semantics (Doron 1991; Schlenker 2003, 2004; Eckardt 2014). In these accounts, (FID) utterances depend on two different contexts: a context of utterance and a context of thought. Personal pronouns and tense markings receive their semantic value from the context of utterance, whereas all remaining indexicals receive their semantic value from the context of thought, i.e., the context of the reported speech event, thereby accounting for the shifting behavior of many indexicals in FID.

An alternative line of work rejects genuine context shift and treats FID as a form of *mixed quotation* (Maier 2015): the reporting speaker quotes the reported speakers utterance/thought while systematically unquoting some material such as tense and pronouns. This unquotation mechanism is governed and mediated by pragmatic principles (Maier 2015, 2017). The effect of multiperspectivity then follows from the selective unquotation of these elements rather than dependence on multiple contexts.

2.3 Re-evaluating the hybridity of FID

Recently, the hybridity claim often made for FID has been reassessed in and refuted by Saure (2025). Instead, Saure argues that the perspectival properties attributed to FID are not a special property that is inherent only to FID, but rather the consequence of more general narrative mechanisms. Consequently, they also apply to the interpretation of perspective-dependent expressions and thus indexicals in ID.

In several experiments on German, Saure (2025) provides compelling experimental evidence that context shift is also licensed in ID, as temporal and local indexical expressions in German ID reports can be interpreted from the reported speaker's perspective rather than the reporting speaker's perspective. According to Saure (2025), three conditions have to be satisfied for this to be possible: i) the ID sentence occurs in a narrative environment, ii) the reported speaker is prominent enough in the discourse to serve as the *perspectival center* and iii) the propositional attitude verb has to introduce an event of speaking or thinking rather than a mental state (Saure 2025: 261f).

In these settings, participants judged ID sentences such as *Er dachte, dass er morgen kommen würde* ('He thought that he would come tomorrow') as natural even though *morgen* is interpreted relative to the reported speaker, not the reporting speaker. FID and ID thus behave alike: both allow perspective-dependent expressions to shift under conditions of narrative embedding and prominence. This empirical convergence motivates Saure's central claim: FID is not a hybrid between DD and ID but the *root-clause equivalent* of ID, since both share their perspectival configuration in narrative contexts. He argues that this parallel between FID and ID has been overlooked in previous research mainly due to the reason that comparably little attention has been paid to ID in general in the formal-semantics literature on speech and thought reports. The few accounts that have investigated ID did so by not putting it into narrative contexts, thereby leaving the striking similarity between ID and FID unnoticed. Put differently, Saure (2025) argues that the alleged hybridity of FID arises from comparing it to an overly restrictive conception of ID. Once ID is recognized as permitting partial context shift, FID no longer requires a special semantic status.

2.4 Zooming in further on ID

The idea that ID allows perspective shift departs from a long-standing consensus. Following Kaplan (1989) and Banfield (1982), most theories maintain that the interpretation of indexicals in ID is fixed by the context of the reporting speaker and therefore abiding the *Fixity Thesis* (cf. Schlenker 2003). Empirical or theoretical discussions have therefore concentrated on the extremes: DD, which straightforwardly instantiates quotation, and FID, which violates Kaplan's restriction and has thus lead to various monstrous analyses of FID.

The result is an uneven research landscape. DD has been thoroughly modeled as direct quotation; FID has inspired an extensive semantics-pragmatics literature on multiperspectivity and narrative viewpoint. ID, however, has often been treated as semantically transparent—as simple propositional embedding under attitude verbs (Hintikka 1969; Kratzer 2006). Given that FID is defined in relation to ID, this neglect is surprising; one cannot explain the mixed behavior of FID without understanding the range of perspectival interpretations available in ID.

Recent experimental work has begun to fill this gap. Anderson (2019) and Saure (2025) show that speakers tolerate non-canonical, shifted readings of temporal adverbs in ID under precisely those conditions that also license viewpoint shift in FID. Similarly, there is research suggesting that non-indexical expressions—such as expressives, appositives, self-pointing gestures, and, as we will discuss in more detail below, face emoji—pattern with some indexical expressions, meaning that they can also be interpreted from the reported speaker’s perspective when embedded in an ID utterance (cf. Harris & Potts 2009; Walter 2025a, 2025b; Walter & Hinterwimmer 2025b). These findings suggest that if one extends the empirical scope beyond indexical expressions, multiperspectivity is possible in ID even outside narrative contexts. Before turning to some background on formal analyses of (face) emoji, we will next briefly turn to discussing the view on quotation we adopt in the present work, namely that it is an iconic act depicting features of its referent, that is, an utterance. This act of depicting is called a *demonstration* (Clark & Gerrig 1990).

3. Quotation

Before summarizing the quotation-as-demonstration framework postulated by Clark & Gerrig (1990), let us briefly return to DD, the report type where quotation plays the most apparent role. Recall that DD is standardly seen as a verbatim reproduction—thus a quote—of another speaker’s utterance. Example (1) is repeated below for convenience.

(1) Yara said, “I will meet you here tomorrow.”

Within the quoted clause, all indexical expressions—*I, you, here, tomorrow*—are interpreted relative to the context of the reported speech event rather than that of the current speaker. DD therefore preserves the perspective of the reported speaker and has traditionally been analyzed as an instance of pure quotation (e.g., Davidson 1979; Partee 1973), meaning that the words enclosed by the quotation marks are merely mentioned, but not used. Consequently, they are commonly assumed to be unable to interact with surrounding linguistic material.

3.1 Traditional approaches to quotation

We would like to briefly sketch the main ideas of two of traditional approaches on quotation found in the formal-semantic literature: the *name-based* approach and the *demonstrative* approach. The name-based approach, which originates from Quine (1940), assumes that quotation marks turn an expression into a name of itself: the quoted string denotes its own linguistic form, such that the quotation in (1) refers to the sentence *I’ll meet you here tomorrow*. Quotation marks, on this view, are essentially operators that shift an expression from use to mention.

The demonstrative approach to quotation has first been articulated in Davidson (1979). It postulates that quotation marks introduce a covert demonstrative pointing to its referent: the quoted utterance. Despite their different theoretical assumptions, a common characteristic of both the name-based and the demonstrative account is that both ascribe a central semantic role to quotation marks, dubbed the *necessity assumption* in De Brabanter (2017, 2023). Essentially, this means that no quotation is possible without quotation marks or any other type of overt

(conventionalized) marking and involves an exact correspondence between the quoted and original form. Consequently, both accounts can only make predictions about overtly marked instances of quotation. This, however, oversimplifies the empirical picture, as it only makes correct predictions for written data, but not for quotation in spontaneous speech.

A large body of empirical and psycholinguistic work shows that quotations in spoken and signed language are often unmarked and far from verbatim. Focusing on spoken language data in this article, Clark & Gerrig (1990) demonstrate that speakers routinely abbreviate or stylize previous utterances, and that listeners nevertheless recognize these as quotations. Such variation is not an exception but an integral property of how people reproduce, that is, quote, linguistic behavior. In fact, interlocutors in a spoken discourse do not even expect a speaker to produce a verbatim repetition of an utterance when quoting someone else's speech (Wade & Clark 1993). This motivates a shift from viewing quotation as an operation tied to quotation marks or a spoken equivalent—whatever this may be²—toward a more liberal notion of quotation viewing it as a mere resemblance relation between the quoted and the original utterance: a demonstration.

3.2 *Quotation as demonstration*

In Clark & Gerrig's (1990) framework, to quote is to demonstrate: rather than describing what someone said, the speaker shows it. The act of quotation involves a form of iconic depiction—the speaker partially reenacts aspects of the original utterance so that the addressee can recognize the resemblance between the two events. Demonstrations therefore convey meaning through similarity. Thus, they are to be contrasted from the ordinary arbitrary mode of communication that is dubbed *description* in Clark & Gerrig (1990).

Demonstrations consist of four components: i) *depictive* aspects, which intentionally mirror features of the original performance (lexical choice, prosody, gesture), ii) *supportive* aspects, which enable the depiction (quotation marks, voice shift, or other cues), iii) *annotative* aspects, commentary or framing material contributed by the current speaker (e.g., *Yara said*), and iv) *incidental* aspects, unintentional by-products of the performance.

This model entails that quotation marks are not semantically necessary; they merely constitute one conventionalized way of signaling that a demonstration is taking place. What matters is that the audience can recognize the speaker's depictive intent—that part of the current utterance is meant to show an earlier one. Clark & Gerrig (1990) posit that each aspect of a demonstration is appropriately marked as such by the speaker so that it can be recognized by the listeners. However, we object at this point that this claim oversimplifies: it is by far not always clear what features of a demonstration are intended to be interpreted as depictive (cf. Walter 2025c for a more thorough discussion). We will set this aside here and leave this matter to future research, although our proposal for ID made in Chapter 5 of this article addresses it to some extent.

The demonstrational view neatly explains why unmarked quotations are pervasive and why speakers often depart from literal reproduction. Note also that in contrast to other theories of quotation discussed above, it extends naturally to multimodal behavior: quotations can

² See De Brabanter (2023) and references cited therein for observations that there may not be a spoken-language equivalent to quotation marks after all, at least not a single feature.

include gestures, facial expressions, or prosodic patterns that resemble those of the original speech event, which is why it will be adopted in the analysis presented in this article.

3.3 *Unmarked mixed quotation*

Further evidence for a demonstration-based account comes from unmarked mixed quotation—cases in which a quoted fragment is syntactically integrated and lacks overt marking. Consider example (4) below.

- (4) Mr. Bailey replied that he hoped he knowed wot o'clock it wos in ginerol.
(Charles Dickens, *Martin Chuzzlewit*, cited from Clark & Gerrig 1990: 791)

This sentence constitutes a clear counterexample to the *necessity claim* and shows that nonstandard dialect can be quoted even in literary indirect discourse. The phenomenon, known as *language shift* (Recanati 2001), involves the narrator adopting the character's speech style within an ID clause. Assuming that the dialectal portion is treated as mixed quotation, (4) and the overtly marked variant in (5) are semantically equivalent.

- (5) Mr. Bailey replied that he hoped he “knowed wot o'clock it wos in ginerol.”
(Kirk-Giannini 2024: 6)

In both sentences, the dialect reflects Mr. Bailey's manner of speaking rather than the narrator's. As Kirk-Giannini (2024) notes, the presence or absence of quotation marks does not affect truth conditions. Similar observations have been made for metalinguistic negation (Potts 2007a), protagonist projection (Stokke 2021), and, as shown above, also for FID (Maier 2015).

These examples show that quotation is best analyzed as an act of depiction that may, but need not, be overtly marked. Once quotation is viewed as demonstration, the difference between marked and unmarked cases reduces to the strength of their supportive cues rather than a categorical grammatical contrast. This also explains why written quotations appear more rigid: orthographic conventions simply replace the prosodic and gestural signals available in speech. It is worth highlighting at this point that the demonstration theory of quotation contrasts sharply with the name-based and demonstrative approaches to quotation in one important respect: while the latter two see quotation as a semantic phenomenon, the demonstrational theory views it as a mode of communication and hence as an instance of language use. Therefore, this approach is, strictly speaking, to be situated within pragmatics, thus posing initial challenges to a formal-semantic analysis of quotation as demonstration. In the following, we will briefly discuss a formal-semantic treatment of quotation proposed in Davidson (2015) that incorporates the idea that quotation is an iconic act of demonstrating.

3.4 *Davidson's (2015) formal implementation*

Davidson (2015) was, to the best of our knowledge, the first to incorporate the insights of Clark & Gerrig (1990) into a framework of quotation, thereby providing a compositional means to integrate demonstrations into truth-conditional semantics. She treats quotation as a semantic phenomenon, however, thus introducing a dedicated type for demonstrations (*d*) into the semantic ontology. Her analysis focuses on a type of quotation thus not mentioned in the

present article, so-called *be like* constructions (cf. (6)), a construction known to also incorporate paralinguistic features as well as gestures into the quote (e.g., Stokke & Ball 2025). In fact, as evidenced by (6b), even non-linguistic behavior can be quoted under quotative *like*.

- (6) a. Bob was like GOBBLING.³
b. My cat was like “feed me”!

(Davidson 2015: 485)

Due to its vividness and flexibility evidenced by the examples in (6), it is especially suited to motivate a demonstrational analysis of quotation. The lexical entry for quotative *like* that she proposes is given in (7).

- (7) $[[\textit{like}]] = \lambda d \lambda e. [\textit{demonstration}(d, e)]$

Crucially, the *demonstration* predicate in (7) holds between a demonstration d and an event e —in other words, d is a demonstration of e —if and only if d reproduces relevant properties of the event e that are relevant in the utterance context. This formal implementation elegantly captures the basic idea that the demonstration, i.e., the quote, and the event it demonstrates need not be identical, but only similar in contextually relevant features.

Yet, also Davidson (2015) holds onto the verbatimness requirement for direct quotes but establishes it with a twist. Consider her lexical entry for *say* that embeds a quote in (8).

- (8) $[[\textit{say}]] = \lambda d \lambda x \lambda e. [\textit{agent}(e, x) \wedge \textit{demonstration}(d, e) \wedge \textit{saying}(e)]$

(Davidson 2015: 488)

The fact that the relevant event is a saying event when embedded under *say* enforces that the relevant properties depicted by the demonstration are the words used in the saying event e . Thus, the verbatimness requirement is mediated by a combination of the lexical semantics of the verb *say* and pragmatics in Davidson’s (2015) account.

While we generally agree with most of Davidson’s (2015) intuitions, especially with the assumption that gestures, facial expression, as well as marked prosody can be part of a demonstration, we will formally spell this out in a different way that we believe is more adequate below in Chapter 5.

3.5 *Interim summary and motivation for the present approach*

Summing up quickly before turning to semantic treatments of (face) emoji, quotation—at least in the framework we adopt in this article—involves two meaning components that are interrelated. First, a descriptive component, which attributes a speech event to a source. Second, quotations involve a depictive meaning component that iconically represents aspects of that event.

Traditional treatments of quotation (Quine 1940; Davidson 1979, et seq.) are only concerned with the descriptive meaning component, meaning that they do not recognize the depictive potential of quotations and consequently fail entirely in capturing it. This can be

³ Small caps indicate an iconic gesture.

attributed to the view still prevalent in formal linguistics that form-meaning mappings, and thus quotation, if taken to be a semantic phenomenon, are arbitrary by nature (de Saussure 1916). Clark & Gerrig's (1990) pragmatic account and Davidson's (2015) integration of their account into a formal-semantic treatment of quotation capture both, explaining why quotations need not be verbatim and why they can extend across modalities. This perspective also supports findings that quotation can occur without formal marking (e.g., Wade & Clark 1993; Potts 2007a; De Brabanter 2017, 2023; Kirk-Giannini 2024). As the demonstration-based account of quotation makes adequate predictions for a wider range of occurrences of quotation in the written, spoken, and signed modality than name-based or demonstrative theories of quotation, we take it to be the superior theoretical approach.

In what follows, we therefore adopt this demonstration-based account of quotation. It allows us to treat instances in which non-indexical elements in ID—such as shifted face emoji—function as unmarked demonstrations, selectively depicting features of the reported speech. Subsequently, we will also extend this analysis to account for shifted local and temporal indexicals in German (Plank 1986; Saure 2025).

4. Emoji

4.1 Introducing emoji

Emoji, small pictorial symbols often used to convey affective meaning, have become a ubiquitous feature of digital communication. Developed in Japan in the late 1990s as successors to text-based emoticons, they were integrated into mobile platforms in the early 2010s; Apple introduced an emoji keyboard in 2011, followed by Android in 2013. Nowadays, they are an essential part of everyday digital written communication. Their success and popularity are most likely due to the absence of prosody and facial as well as manual gesture in digital communication, where emoji simply function as multimodal cues adding mostly affective or pragmatic meaning to the message they accompany.⁴ There are various subclasses of emoji. We will focus on so-called *face emoji* in this article, that is, stylized depictions of facial expressions such as smiling (e.g., 😊) or frowning (e.g., ☹️). They play a central role in conveying an author's emotional state toward the accompanying text or, more broadly, the situation described by that text (Grosz et al. 2023). This underlines the above-made claim regarding their popularity, as they provide a visual substitute for facial expression, and partly also prosodic cues that are typically associated with the emotion expressed by that facial expression.

From a theoretical perspective, emoji occupy an interesting position between linguistic and non-linguistic systems of meaning. Like words, they can combine compositionally with text—especially when they replace text and are thus used as *pro-text* emoji (cf. Pierini 2021)—and show conventionalized patterns of use; yet like gestures or pictures, they are at least partly

⁴ Note that we remain agnostic at this point with regard to the question whether emoji actually add meaning *compositionally* to the sentence they accompany or whether they simply *comment on* the proposition expressed by the message they accompany and their contribution is thus mediated by discourse structure, as we are not too concerned with the core semantic and pragmatic functions of emoji in this paper. Note that for our purposes, this discussion is mostly irrelevant.

iconic, deriving aspects of their meaning from resemblance rather than from arbitrary convention. In this respect, they are often compared to both expressives in the sense of Potts (2005) (Grosz et al. 2023) and iconic gestures (Pierini 2021). A consistent observation across these lines of research is that emoji typically contribute their meaning *not-at-issue*: they add backgrounded information that cannot be directly denied in discourse, thus not being ‘on the table’ for discussion, to use the metaphor of Farkas & Bruce (2010). They thus pattern with other types of iconic enrichments like iconic gestures (Ebert & Ebert 2014; Schlenker 2018; Ebert et al. 2020) and ideophones (Henderson 2016; Barnes et al. 2022; Barnes & Ebert 2023; Barnes 2024) in this respect. Moreover, their interpretation is by default *author-oriented*—they express the affective state or stance of the message’s producer rather than that of a mentioned protagonist (Kaiser & Grosz 2021). Before returning to their perspectival properties in more detail and challenging claims made in previous research, we will briefly discuss two complementary theoretical accounts that have been proposed to capture the properties of emoji just discussed: the lexicalist approach (Grosz et al. 2023) and the pictorial approach defended in Maier (2023).

4.2 Semantic accounts of (face) emoji

To our knowledge, two different semantic accounts of emoji have been proposed in recent years. The lexicalist account, as developed in Grosz et al. (2023) based on several experimental studies reported in earlier work (e.g., Kaiser & Grosz 2021), treats emoji as having a fixed, conventionalized meaning that is similar to expressives like *damn* (Potts 2005; cf. also Grosz 2025 for a more comprehensive survey on the expressive potential of especially face emoji). Thus, since they are treated on the same level as words, they are, from a semiotic point of view, symbols. The pictorial account proposed in Maier (2023), by contrast, argues that emoji are “used to depict ‘what the world looks like’” (Maier 2023: 303), hence they are treated on a par with pictures, having a completely iconic semantics. Building on this, emoji are nothing more than a highly schematic representation of the author’s facial expression while typing the message.

Starting with the view defended in Grosz et al. (2023), face emoji are conventional lexical items that contribute *use-conditional* meaning in the sense of Gutzmann (2015)—parallel to expressions such as expressive adjectives like *damn* or emotive markers like *alas* (cf. Rett 2021) in spoken language. Each emoji specifies a fixed emotional stance (e.g., happy or unhappy) toward the proposition expressed by the accompanying text. Formally, a (face) emoji denotes a relation between the author of the message, a target proposition, and a discourse value representing what the author desires or approves of. A smiling face, for example, denotes the set of worlds in which the author is happy about how the target proposition promotes their discourse value, while a frowning face denotes the opposite relation. Because the first argument of this denotation is the author variable of the discourse context, the affective attitude is automatically evaluated with respect to the author. In this way, the model derives the observed author-orientation compositionally, rather than stipulating it as a pragmatic default. The not-at-issue status of emoji content follows from their expressive type: the affective comment is added on a separate dimension of meaning, much like Potts’s (2005) conventional implicatures, and cannot be straightforwardly denied. Nevertheless, the framework leaves room for occasional perspective shifts, as the author variable can, in principle, be re-anchored when another participant’s perspective becomes highly prominent.

Crucially, however, shifted interpretations of face emoji in ID are illicit according to Kaiser & Grosz (2021).

Maier (2023), by contrast, rejects the view that emoji are lexical expressions at all. He argues that they are pictures: schematic, stylized depictions governed by a pictorial semantics of geometric projection (Greenberg 2013, 2021; Abusch 2013, 2020). On this view, an emoji's basic content is minimal—it represents the set of worlds that, from some viewpoint, would look like the image displayed. This model accounts for the iconic, non-arbitrary nature of emoji and for their flexibility across contexts. Importantly, Maier distinguishes between truth-conditional and use-conditional pictorial content. Emoji depicting objects or scenes (e.g., 🚗 or 🎁) contribute truth-conditional information about what the world looks like, whereas face and hand emoji depict what the utterance context looks like—specifically, what the speaker's face or gesture looks like while producing the message. Their expressivity, captured by them adding use-conditional information, thus arises from the depicted viewpoint: the image shows the speaker's own facial expression and thereby conveys the author's emotional attitude in an iconic, i.e., non-symbolic, manner. Author-orientation follows automatically from the canonical viewpoint Maier (2023) proposes for face emoji, since it corresponds to an addressee looking at the face of the author. Thus, in order to yield a shifted interpretation of a face emoji, one would have to shift the viewpoint away from the canonical viewpoint, which is a marked move in pragmatic terms and thus only rarely performed. This explains the hard-wired, default author-orientation of face emoji predicted for expressives in general (cf. Potts 2007b).

While both theories capture the affective and multimodal nature of emoji, they differ in where they locate their meaning. Grosz et al. (2023) adopt the view that emoji are part of the linguistic system, with conventional expressive denotations anchored to the author. Maier (2023), by contrast, treats them as visual depictions whose meaning derives from resemblance and viewpoint. The former view thus emphasizes compositional integration and conventionalization, the latter the iconic core of them. Despite their differences, both converge on the view that face emoji contribute non-truth-conditional, author-oriented meaning, consequently predicting that shifted interpretations of them are not possible in ID. In the following, subsection, we will present data suggesting that this prediction is untenable from an empirical perspective.

4.3 Shifted face emoji in ID?

Previous research has generally assumed that face emoji are rigidly author-oriented, that is, that their affective contribution invariably reflects the emotional stance of the message's producer rather than that of a reported protagonist. Grosz et al. (2023: 909) explicitly argue that face emoji cannot undergo perspective shift in ID. It is mostly in sentences containing either object-experiencer (cf. (9a) for the verb *to annoy*) or subject-experiencer (cf. (9b) for the verb *to admire*) psych verbs where face emoji can receive non-author-oriented interpretations.

- (9) a. Richie annoyed Adrian. 😞
b. Daniel admires Aaron. 😊

(Grosz 2025: 790)

In Walter & Hinterwimmer (2025b), we scrutinize whether the assumption that face emoji can never shift in ID is empirically tenable. Intuitively, however, it seems that face emoji can receive non-author-oriented interpretations in ID after all:

(10) Peter complained that he had to pay the bill for the whole group again. 🙄

In (10), we argue that the most plausible interpretation of the pouting face emoji (🙄) is one from the perspective of Peter, not from the perspective of the author of the message. This intuition was confirmed in a rating study we report in Walter & Hinterwimmer (2025b). In this study, 30 native speakers of German rated 24 items similar to (11) (the original items were in German) on a 7-point Likert scale. Crucially, the only plausible interpretation of the emoji was from the perspective of the matrix subject, i.e., the reported speaker. We manipulated perspective prominence on the speech level (author vs. matrix subject). A sample item that has been translated from German is given in (11).

- (11) a. Pia wondered why her best friend Anna, that damned lush, had poured her too much wine again last night, even though she couldn't handle it. 🤢
b. Pia wondered why her best friend Anna, who just wanted to get rid of the last drops in the bottle, had poured her too much wine again last night, even though she couldn't handle it. 😬

(Walter & Hinterwimmer 2025b: 1670)

The content of the appositive *that damned lush* in (11a) clearly reflects Pia's—not the author's—negative attitude toward Anna, making her perspective the more prominent one in this example. We therefore hypothesized that the nauseated face emoji can be interpreted from Pia's perspective, i.e., expressing her disgust and hungoverness. Conversely, in (11b) the content of the appositive *who just wanted to get rid of the last drops in the bottle* highlights the perspective of the message's author, inviting an interpretation of the emoji that comments on the event from their perspective. However, since there is no reason why the author should feel disgusted about the situation, sentences of this type should receive comparably lower ratings than sentences of the type in (11a).

The results confirm that emoji can be interpreted from the reported speaker's point of view rather than from the author's (the reporting speaker), especially when their perspective is made prominent in the surrounding discourse. Recall that this is in line with Saure's (2025) findings for shifted interpretations of local and temporal indexicals in German ID utterances.

We argue that shifted face emoji in ID are quoted facial expressions from the reported speaker. Crucially, these quoted facial expressions were instances of unmarked mixed quotation (cf. Kirk-Giannini 2024). To model this, we propose a more fine-grained version of the formal analysis of ID proposed in Walter (2025b). We will then also show that our analysis is also able to account for the cases of shifted indexicals in German reported in Saure (2025).

5. Formal proposal

As mentioned in the last chapter already, in §5.1 we propose a formal analysis of shifted face emoji in ID treating them as cases of unmarked mixed quotation (e.g., Kirk-Giannini 2024). We will then show that this analysis naturally extends to shifted indexicals in stretches of ID in German (Plank 1986; Saure 2025). Before turning to our proposal, however, we need to make some theoretical considerations on what such an analysis should look like and why previous analyses fail to straightforwardly capture the shifting behavior of ID.

One of our core assumptions is that ID utterances can contribute their meaning in multiple dimensions of meaning if they contain quoted elements (cf. Potts 2007a, for this assumption in other quotational structures). They make a foregrounded, that is, *at-issue* contribution that roughly paraphrases as there existing a reported speech event and that the content of this speech event is reported in the ID utterance. The second meaning contribution is *not-at-issue* and thus backgrounded. We argue that this secondary meaning contribution is iconic in that it selectively depicts certain aspects of the reported speech event. In other words, it is a demonstration in the sense of Clark & Gerrig (1990). To model this, we will implement the formal-semantic analysis of quotation as demonstration (Davidson 2015) into our framework, thereby treating demonstrations as part of the semantic ontology. This means that shifted face emoji and shifted temporal and local indexicals will be analyzed as demonstrations in our account.

To illustrate the multidimensionality of instances of mixed quotation in ID utterances, we will first give some background on not-at-issue content. The concept can be introduced most easily by contrasting it with what is called *at-issue content* in Potts (2005), which essentially is the main point of an utterance or, adopting a more discourse-oriented view, the point the speaker wants to steer the conversation toward. Not-at-issue content, then, is the exact opposite of at-issue content: it is considered to be discourse-peripheral, merely commenting on the main part—the at-issue part—of the utterance, or an aside that the speaker does not wish to discuss any further. Appositives have been argued to contribute not-at-issue content by default (Potts 2005). The peripheral status attributed to not-at-issue content is often argued to result in its having special semantic and discourse properties: for example, it has been shown to escape the scope of semantic operators such as negation, a property commonly known as *projectivity* (e.g., Langendoen & Savin 1971; Heim 1983; Geurts 1999; Potts 2005). From a theoretical point of view, not-at-issue content has been argued to not carry information that is relevant to the current *question under discussion* (QUD, henceforth, see Simons et al. 2010)⁵, thereby explaining its peripheral semantic as well as discourse status. Furthermore, and probably also due to its peripheral status, it has been noted that it cannot be directly denied in discourse (cf. B₁ in (12)). This means that in order to disagree with content of a not-at-issue proposition, one must use different means than ordinary sentential negation. For example, one could use a discourse-interrupting protest such as *Hey, wait a minute!* before targeting the not-at-issue content (Shanon 1976; von Stechow 2004). Originally proposed for presuppositional content only, this test has later been extended to diagnose not-at-issue content, in general (Pearson 2010). Alternatively, as illustrated by B₂'s response in (12), one can target not-at-issue propositions by first assenting with the at-issue content followed by an adversative

⁵ Due to reasons of space, we cannot provide a detailed theoretical overview on different theoretical conceptualizations of not-at-issueness in this article. The interested reader is referred to Koev (2018) for a comprehensive overview. We will adopt a QUD-based approach to at-issueness throughout this article.

continuation (diagnostic #1c in Tonhauser 2012). A's utterance in (12) is taken from Syrett & Koev (2015).

- (12) A: My friend Sophie, a classical violinist, performed a piece by Mozart.
B₁: #No, she is a trumpet player.
B₂: Yes, true, but she's not a classical violinist.
B₃: #Yes, true, but she didn't perform a piece by Mozart.

First and foremost, the example illustrates the general non-deniability of not-at-issue content (B₁). The utterance of B₂ in (12) shows how Tonhauser's (2012) diagnostic #1c works for not-at-issue content. Finally, the response of B₃ in (12) illustrates that this diagnostic really works only to target not-at-issue content, as their response targets the at-issue content contributed by the main clause of A's utterance, thus yielding infelicity.

To get a clearer understanding of the at-issue/not-at-issue divide in ID utterances containing shifted face emoji, consider the next example, where diagnostic #1c of Tonhauser (2012) is applied:

- (13) A: Peter complained that he had to pay for the whole group again. 😡
B₁: Yes, true, but he didn't make such an angry face when saying it.
B₂: #Yes, true, but he didn't complain about it.
B₃: #Yes, true, but it wasn't Peter who said that.
B₄: #Yes, true, but he didn't complain about having to pay for the whole group again.

The response of B₁, targeting the content of the face emoji of A's message, is felicitous. This suggests that the emoji's contribution is not-at-issue, meaning that it does not address the QUD. Thus, we assume that it is part of the quotation, as we will show in more detail below. By contrast, the responses of B₂ and B₃ come out as infelicitous, therefore suggesting that the facts that the reported speech event was an event of complaining and that Peter was the agent of this complaining event form part of the at-issue content. Finally, the infelicity of B₄'s response also suggests that the content of the proposition embedded under the verb of saying—to *complain* in the case of (13)—contributes meaning in the at-issue dimension. Thus, the type of the reported speech event, fixing who the agent of this speech event was, and the content of that speech event all form part of the descriptive component of an ID utterance. To keep track of this divide, we will introduce propositional variables into our formal apparatus, where all at-issue content will be indexed with the variable *p*, whereas all not-at-issue content will be indexed with *p** (cf. Ebert et al. 2020).

For the analysis, we follow suggestions made by Angelika Kratzer (2006, 2016) that propositional attitude verbs are transitive verbs that take a special kind of direct object, namely belief objects, as their argument. This analysis has been extended to quotation in Maier (2020). Kratzer thereby refuses the traditional analysis of propositional attitude verbs treating them as universal modal quantifiers (Hintikka 1969). She formally spells this out in a neo-Davidsonian event semantics framework. For our analysis, we will adopt this view, consequently also situating it in event semantics. Furthermore, we will assume that speech events have a special property that sets them apart from all other types of events. Besides the properties they share with other types of events, such as having an agent, a theme, etc., we assume that they additionally have a form and a content property (Maier 2017). The content of a speech event

is taken to be its propositional content. In an analysis for quotation, Maier (2018: 269) makes use of this distinction between form and content, assigning the following semantics to quotation marks:

$$(14) \quad \text{“...”} \sim \lambda q \lambda e. [\text{form}(e, q)]$$

Under this treatment, quotation marks are a function taking a (speech) event and the quoted string of words as its arguments, imposing that both need to have the same form. Put differently, it imposes the verbatimness constraint. Consider the following example for illustration, where this proposed semantics is applied to a stretch of DD:

- (15) a. Daniel said, “I’m hungry”.
b. $\exists e [\text{say}(e) \wedge \text{agent}(e, \text{daniel}) \wedge \text{form}(e, \ulcorner \text{I’m hungry} \urcorner)]$

In Maier’s framework the sentence in (15a) is thus true, iff there exists a saying event with Daniel as its agent and the form of that speech event is $\ulcorner \text{I’m hungry} \urcorner$. The Quine corners indicate that something is of the utterance type u in the sense of Potts (2007b). Note also that Maier (2018) follows many other researchers in quotation by assigning a strong semantic function to quotation marks, thereby hardwiring the aforementioned necessity claim (De Brabanter 2023) into his account.

For an ID utterance as in (16a), Maier (2017) proposes a semantics along the lines of (16b). This amounts to the lexical entry for *to say* in its use as a propositional attitude verb provided in (16c).

- (16) a. Daniel said that he was hungry.
b. $\exists e [\text{say}(e) \wedge \text{agent}(e, \text{daniel}) \wedge \text{content}(e, \llbracket \text{he was hungry} \rrbracket)]$
c. $\llbracket \text{say} \rrbracket = \lambda p \lambda x \lambda e. [\text{say}(e) \wedge \text{agent}(e, x) \wedge \text{content}(e, p)]$

In the proposal of Maier (2017, 2018), the main difference between ID and DD thus boils down to whether the speaker remains faithful to the content (ID) or to the form (DD) of the utterance.

We agree in large parts with Maier’s proposals for ID and DD. However, a major shortcoming of this account, in particular of the lexical entry proposed for a verb of saying, is that it does not make any predictions regarding shifted temporal and local indexicals (Plank 1986; Anderson 2019; Saure 2025) and face emoji (Walter & Hinterwimmer 2025b) in ID. We argue that this can be fixed rather easily: one simply needs to propose a lexical entry for verbs of saying that also takes into account the form of a speech event.

5.1 Accounting for shifted face emoji

To formalize the intuitions above, we need to do two things: first, to account for the proposal that mixed quotation is at play in ID, we need to integrate into our account the idea that the form of (parts of) the speech event play a role in determining the truth conditions of an ID utterance, as well. Second, since we view quotation as a multimodal phenomenon (Stec 2012, 2016), we see it as an instance of the mode of communication called demonstration in Clark & Gerrig (1990). We will therefore integrate the formal analysis of Davidson (2015) into our proposal, i.e., we will assume that demonstrations are part of the semantic ontology and are

therefore of the dedicated semantic type d . We achieve the first goal by assuming that a subevent of the reported speech event e and the demonstration of this speech event d need to be similar in contextually relevant aspects. To model this, we draw from work in gesture semantics and implement the two-place, not-at-issue similarity predicate SIM_{p^*} proposed in Ebert et al. (2020) into our framework (cf. also Umbach & Gust 2014).

Before turning to the formal proposal, we would like to stress that we take gestures, facial expressions, and prosodic patterns to be subevents of a speech event. When any of these elements receives a shifted interpretation—or, in our terms, an interpretation as a quoted element—it functions as a demonstration that iconically resembles a subevent of the reported utterance (e'') (cf. Schlenker & Lamberton 2024, for a similar analysis of classifier constructions in sign languages). As we take face emoji themselves to be depictions of facial expressions, they can also saturate the demonstration argument in our proposal. The content relation holds between the embedded proposition and the remaining part of the speech event (e').

With these assumptions in place, we are now ready to provide a lexical entry for a verb of saying. Consider again example (10) above, repeated in (17) for convenience.

(17) Peter complained that he had to pay the bill for the whole group again. 🙄

For the verb of saying *to complain*, we propose the following lexical entry that takes into account all our assumptions from above:

(18) $\llbracket \text{complain} \rrbracket = \lambda p \lambda d \lambda x \lambda e. [\text{complain}_p(e) \wedge \text{agent}_p(e, x) \wedge \exists e' \subseteq e \exists e'' \subseteq e' [\text{content}_p(e', p') \wedge \text{SIM}_{p^*}(\text{form}(e''), d)]]$

Thus, after λ -conversion and existential closure (Heim 1982), this leaves us with the truth conditions for (17) as stated below.

(19) $\llbracket (17) \rrbracket = \exists e [\text{complain}_p(e) \wedge \text{agent}_p(e, \text{peter}) \wedge \exists e' \subseteq e \exists e'' \subseteq e' [\text{content}_p(e', \llbracket \text{he had to pay for the whole group again} \rrbracket) \wedge \text{SIM}_{p^*}(\text{form}(e''), \text{🙄})]]$

According to (19), the sentence in (17) is true iff there exists a complaining event whose agent is Peter, and the content of a subevent of this complaining corresponds to the proposition embedded under *to complain*. In addition, there must be another subevent—the facial expression accompanying Peter’s speech act—whose form is similar to the pouting emoji occurring in the ID report. In this configuration, the emoji saturates the demonstration argument associated with *to complain*, meaning that it constitutes the only quoted element in (19). The propositional variables mark that only the final conjunct—the similarity predicate imposing the form constraint—contributes not-at-issue meaning, while all other parts belong to the at-issue dimension, thus reflecting the judgments obtained from (13) above.

5.2 Extension to shifted indexicals in German

As already mentioned in §2.3 and §2.4 above, Saure (2025) has provided clear empirical evidence that temporal and local indexicals can receive shifted interpretations, on which they

are interpreted from the perspective of the subject of the propositional attitude verb rather than from the speaker’s or narrator’s perspective, not only in FID, but also in ID. Recall, however, that the availability of such readings depends on the following three conditions to be fulfilled: i) The ID sentence occurs in a narrative environment, ii) the reported speaker is prominent enough in the discourse to serve as the *perspectival center*, and iii) the propositional attitude verb has to introduce an event of speaking or thinking rather than a mental state. Now, the third condition makes an analysis of shifted readings of local and temporal indexicals in ID in terms of mixed quotation quite attractive since it can directly be derived from such an account: For speech or thought events, it is natural to assume that they have a form that can be (partially) quoted in addition to a content, while the assumption that abstract mental states have a form is rather counterintuitive. In light of these considerations, let us therefore end this paper by sketching an account of shifted readings of local and temporal indexicals in ID that assumes propositional attitude verbs introducing events of speaking or thinking to have the lexical entries proposed in the preceding chapter.

Recall that on our account such verbs introduce two subevents of the respective event of speaking or thinking that they denote and take one argument which provides the content of the first subevent and another argument which provides a demonstration that is similar (in relevant respects) to the form of the second subevent. With these assumptions in place let us now see how our analysis can account for the shifted reading of the temporal indexical *yesterday* in (20a), assuming the lexical entry for *say* given in (20b).

- (20) a. Martha said that she had finished her thesis yesterday.
b. $[[\mathbf{say}]] = \lambda p \lambda d \lambda x \lambda e. [\text{say}_p(e) \wedge \text{agent}_p(e, x) \wedge \exists e' \subseteq e \exists e'' \subseteq e [\text{content}_p(e', p') \wedge \text{SIM}_{p^*}(\text{form}(e''), d)]]$

Now, in order to receive a shifted interpretation, *yesterday* has to saturate the demonstration argument of *say* rather than be a part of the first argument, which provides the content of the first subevent. Since temporal as well as local adverbs are optional elements by definition, it is unproblematic to assume that they occupy a peripheral position at LF at the latest (potentially via covert movement), where they can saturate the demonstration argument of *say* independently of the rest of the embedded clause, which saturates the propositional argument of *say*. Note that the rest of the embedded clause always denotes a complete proposition in itself, in virtue of local and temporal indexicals being optional. After λ -conversion and existential closure, the sentence in (20a) is thus interpreted as given in (21):

- (21) $\exists e [\text{say}_p(e) \wedge \text{agent}_p(e, \text{peter}) \wedge \exists e' \subseteq e \exists e'' \subseteq e [\text{content}_p(e', [\mathbf{she had finished her thesis}]) \wedge \text{SIM}_{p^*}(\text{form}(e''), \text{yesterday})]]$

The adverb *yesterday* is thus interpreted as being quoted from Martha’s original utterance, while the rest of the clause provides (the remaining part of) the content rather than necessarily being similar in its form to that utterance.

Now, a clear advantage of the account just sketched is that it does not have to stipulate that only indexical adverbs, but not first- and second-person pronouns can receive shifted readings in ID. Rather, this directly follows from the former being optional elements that are not necessary for the clause containing them to denote a complete proposition, while the latter

cannot be separated from the clause containing them without that clause ceasing to denote a complete proposition. Note, however, that it follows from this assumption that shifted readings of indexicals in languages such as Amharic and Zazaki, which also allow first- and second-person pronouns to receive such readings (Schlenker 2003; Anand & Nevins 2004; Anand 2006; Pearson 2012; Deal 2020), have to come about by a different mechanism. Given the easy availability of such readings in these languages that seems to be entirely unconstrained by pragmatic factors, however (see Deal 2020 for discussion), this seems to be a natural assumption, anyway. Let us return to the first two conditions assumed by Saure (2025) to be prerequisites for the availability of shifted readings of local and temporal indexicals. Regarding the second condition, it fits naturally with the assumption that such readings come about in essentially the same way as shifted readings of face emoji, since we observe the same constraint for face emoji (see §5.1 above and the experimental evidence discussed in detail in Walter & Hinterwimmer 2025b). Regarding the first condition, there seems to be a difference between the two cases that requires further investigation, since it is not obvious that a narrative environment is required for the availability of shifted readings of face emoji.

Let us finally turn to a prediction that our account makes for shifted readings of local and temporal indexicals that might at first sight seem to be problematic. Recall that we assume the part of the respective proposition stating the similarity of the demonstrated element and the form of a subevent of the reported speech or thought event to be not at-issue. For local and temporal indexical adverbials, this might seem counterintuitive at first. After all, it may well be that the information they convey is part of the at-issue content of the respective speech or thought report, as shown by the felicity of the discourse in (22):

- (22) A: Martha said that she had finished her thesis yesterday.
B: No, she said that she had finished her thesis the day before yesterday.

Note, however, that if there is a subevent of the reported speech event whose form is *yesterday*, the content of that subevent (namely, the meaning of *yesterday*) is necessarily also part of the content of the reported event. Since the addressee can be assumed to make that inference as well, we tentatively suggest that it is possible to treat the content of shifted indexical local or temporal adverbs as at-issue content despite the information regarding their form being not at-issue. Clearly, this issue requires a more detailed investigation, which goes beyond the scope of this paper, however, and which we therefore have to leave for future research.

6. Discussion and conclusion

In this article, we have argued for an analysis of ID that permits mixed quotation. This analysis not only accounts for shifted interpretations of face emoji but also for shifted interpretations of local and temporal indexical adverbials in languages such as German and English—languages traditionally categorized as not allowing indexical shift. However, this categorization appears empirically inadequate: increasing evidence shows that even in these languages shifted readings of indexical expressions are possible, at least under the right pragmatic circumstances (e.g., Plank 1986; Anderson 2019; Saure 2025).

To capture both the data on shifted face emoji reported in Walter & Hinterwimmer (2025b) and the data on shifted indexicals in non-shifting languages, we adopted the quotation-

as-demonstration account originally proposed by Clark & Gerrig (1990) and adapted to formal semantics by Davidson (2015). This framework treats quotation as an iconic act depicting certain features of its referent, i.e., the reported speech event. Crucially, because it allows that aspects of speech events beyond words can be quoted, it can straightforwardly model shifted interpretations of face emoji by treating them as quoted facial expressions of the reported speaker. Our analysis is couched in an event-semantic framework in which speech events have both form and content components (Maier 2018). We model quotation via a similarity predicate SIM, which relates a subevent of the reported speech event to the demonstration argument posited for verbs of saying and thinking. We further stipulate that only *optional* material can saturate this demonstration argument, which correctly predicts that only indexical adverbials—elements not required for the truth conditions of the sentence—can receive shifted interpretations in traditionally non-shifting languages.

We therefore propose a crosslinguistic divide in how indexical shift is licensed. For languages such as Amharic, which allow systematic indexical shift independent of pragmatic factors, shifty-operator approaches (e.g., Anand & Nevins 2004; Anand 2006; Deal 2020) remain the most adequate. For languages such as German and English, traditionally thought to disallow indexical shift—a claim we have shown to be empirically untenable—we argue that a mixed-quotation analysis best captures the observed patterns. Since our account predicts that only optional material can be quoted in ID, it naturally explains why only indexical adverbials can shift in these languages. Although a comprehensive crosslinguistic investigation of indexical shift in non-shifting languages remains a task for future research, the available data suggest that the simple two-way typology between shifting and non-shifting languages is empirically inadequate. Instead, it appears that many—perhaps all—languages permit indexical shift to some degree; what varies crosslinguistically is the ease with which it is licensed. A promising avenue for future research would thus be to re-assess the shifting potential of languages such as English or Spanish, which have been assumed to be non-shifting languages in the traditional dichotomy that only differentiates between shifting and non-shifting languages (e.g., Anand & Nevins 2004).

Our proposal also raises several theoretical questions. First, the account predicts that only elements in peripheral positions can saturate the demonstration variable and thus be quotable. We therefore predict that shifted indexical adverbials must move to a peripheral position at least at LF—an assumption that needs to be worked out in more detail in future work. Second, we have treated the similarity predicate as contributing not-at-issue meaning by default. As illustrated by example (22), this assumption may not be entirely appropriate for shifted indexicals and warrants further investigation. Moreover, our analysis currently assumes that quotation is a constant feature of ID—that verbs of saying and thinking obligatorily include a demonstration argument. While we have argued that quotation is more persistent in ID than commonly assumed, it may nonetheless be too strong to posit a demonstration variable as part of every lexical entry for verbs of saying and thinking. A plausible alternative is to assume that verbs of saying and thinking normally lack a demonstration argument, but that a type-shifting operation can introduce one when pragmatic factors make a quotational reading sufficiently salient. Developing such an approach remains a task for future research.

References

- Abusch, Dorit. 2013. Applying discourse semantics and pragmatics to co-reference in picture sequences. In Chemla, Emmanuel & Homer, Vincent & Winterstein, Grégoire (eds.), *Proceedings of Sinn und Bedeutung (SuB) 17*, 9–25. Paris: École Normale Supérieure Paris.
- Abusch, Dorit. 2020. Possible worlds semantics for pictures. In Gutzmann, Daniel & Matthewson, Lisa & Meier, Cécile & Rullmann, Hotze & Zimmermann, Thomas Ede (eds.), *The Wiley Blackwell companion to semantics*, 1–20. Hoboken, NJ: Wiley Blackwell.
- Anand, Pranav & Nevins, Andrew. 2004. Shifty operators in changing contexts. In Young, Robert B. (ed.), *Proceedings of Semantics and Linguistic Theory (SALT) 14*, 20–37. Ithaca, NY: CLC Publications.
- Anand, Pranav. 2006. *De de se*. Cambridge, MA: MIT. (Doctoral dissertation.)
- Anderson, Carolyn Jane. 2019. Tomorrow isn't always a day away. In Espinal, M. Teresa & Castroviejo, Elena & Leonetti, Manuel & McNally, Louise & Real-Puigdollers, Cristina (eds.), *Proceedings of Sinn und Bedeutung (SuB) 23*, 37–56. Barcelona: Universitat Autònoma de Barcelona.
- Banfield, Ann. 1982. *Unspeakable sentences: Narration and representation in the language of fiction*. Boston, MA: Routledge.
- Barnes, Kathryn. 2024. *The semantics and pragmatics of German ideophones: Warum geht der Frosch plitsch plitsch die Treppe hoch?*. Frankfurt/Main: Goethe University Frankfurt. (Doctoral dissertation.)
- Barnes, Kathryn & Ebert, Cornelia. 2023. The information status of iconic enrichments: Modelling gradient at-issueness. *Theoretical Linguistics* 49(3-4). 167–223.
- Barnes, Kathryn & Ebert, Cornelia & Hörnig, Robin & Stender, Theresa. 2022. The at-issue status of ideophones in German: An experimental approach. *Glossa: A Journal of General Linguistics* 7(1). 1–39.
- Clark, Herbert H. & Gerrig, Richard J. 1990. Quotations as demonstrations. *Language* 66(4). 764–805.
- Davidson, Donald. 1979. Quotation. *Theory and Decision* 11(1). 27–40.
- Davidson, Kathryn. 2015. Quotation, demonstration, and iconicity. *Linguistics and Philosophy* 38(6). 477–520.
- Deal, Amy Rose. 2017. *Shifty asymmetries: Universals and variation in shifty indexicality*. Unpublished.
- Deal, Amy Rose. 2020. *A theory of indexical shift: Meaning, grammar, and crosslinguistic variation*. Cambridge/London: MIT Press.
- De Brabanter, Phillipe. 2017. Why quotation is not a semantic phenomenon, and why it calls for a pragmatic theory. In Depraetere, Ilse & Salkie, Raphael (eds.): *Semantics and pragmatics: Drawing a line*, 227–254. Cham: Springer.
- De Brabanter, Phillipe. 2023. Quotation does not need marks of quotation. *Linguistics* 61(2). 285–316.
- De Saussure, Ferdinand. 1916. *Cours de linguistique générale*. Paris: Payot.
- Doron, Edit. 1991. Point of view as a factor of content. In Moore, Steven & Wyner, Adam Zachary (eds.): *Proceedings of Semantics and Linguistic Theory (SALT) 1*, 51–64. Ithaca, NY: Cornell University Working Papers in Linguistics.
- Ebert, Christian & Ebert, Cornelia & Hörnig, Robin. 2020. Demonstratives as dimension shifters. In Franke, Michael & Kompa, Nikola & Liu, Mingya & Mueller, Jutta L. & Schwab, Juliane (eds.), *Proceedings of Sinn und Bedeutung (SuB) 24*, 161–178. Osnabrück: University of Osnabrück.
- Ebert, Cornelia & Ebert, Christian. 2014. Gestures, demonstratives, and the attributive/referential distinction. Talk presented at *Semantics and Philosophy in Europe*. Berlin.

- Eckardt, Regine. 2014. *The semantics of free indirect discourse: How texts allow to mind-read and eavesdrop*. Leiden: Brill.
- Farkas, Donka F. & Bruce, Kim B. 2010. On reacting to assertions and polar questions. *Journal of Semantics* 27(1). 81–118.
- Geurts, Bart. 1999. *Presuppositions and pronouns*. Oxford: Elsevier.
- Greenberg, Gabriel. 2013. Beyond resemblance. *Philosophical Review* 122(2). 215–287.
- Greenberg, Gabriel. 2021. Semantics of pictorial space. *Review of Philosophy and Psychology* 12(4). 847–887.
- Grosz, Patrick G. 2025. Expressivity and emojis. In Gutzmann, Daniel & Turgay, Katharina (eds.), *The Oxford handbook of expressivity*, 776–801. Oxford: Oxford University Press.
- Grosz, Patrick G. & Kaiser, Elsi & Pierini, Francesco. 2021. Discourse anaphoricity and first-person indexicality in emoji resolution. In Grosz, Patrick G. & Martí, Luisa & Pearson, Hazel & Sudo, Yasutada & Zobel, Sarah (eds.), *Proceedings of Sinn und Bedeutung (SuB) 25*, 340–357. London: University College London.
- Grosz, Patrick G. & Greenberg, Gabriel & De Leon, Christian & Kaiser, Elsi. 2023. A semantics of face emoji in discourse. *Linguistics and Philosophy* 46(4). 905–957.
- Gutzmann, Daniel. 2015. *Use-conditional meaning: Studies in multidimensional semantics*. Oxford: Oxford University Press.
- Harris, Jesse A. & Potts, Christopher. 2009. Perspective-shifting with appositives and expressives. *Linguistics and Philosophy* 32(6). 523–552.
- Henderson, Robert. 2016. A demonstration-based account of (pluractional) ideophones. In Moroney, Mary & Little, Carol-Rose & Collard, Jacob & Burgdorf, Dan (eds.), *Proceedings of Semantics and Linguistic Theory (SALT) 26*, 664–683. Austin, TX: LSA Publishing.
- Heim, Irene. 1982. *The semantics of definite and indefinite noun phrases*. Amherst, MA: University of Massachusetts Amherst. (Doctoral dissertation.)
- Heim, Irene. 1983. On the projection problem for presuppositions. In Flickinger, Daniel P. (ed.), *Proceedings of the West Coast Conference on Formal Linguistics (WCCFL) 2*, 114–125. Stanford, CA: Stanford Linguistics Association.
- Hintikka, Kaarlo Jaakko Juhani. 1969. Semantics for propositional attitudes. In Hockney, Donald J. (ed.), *Philosophical Logic*, 21–45. Dordrecht: Reidel.
- Kaiser, Elsi & Grosz, Patrick G. 2021. Anaphoricity in emoji: An experimental investigation of face and non-face emoji. In Farrell, Patrick (ed.), *Proceedings of the Linguistic Society of America (LSA) 6*, 1009–1023. Lexington, KY: LSA Publishing.
- Kaplan, David. 1989. Demonstratives. In Almog, Joseph & Perry, John & Wettstein, Howard (eds.), *Themes from Kaplan*, 565–614. Oxford: Oxford University Press.
- Kirk-Giannini, Cameron Domenico. 2024. Covert mixed quotation. *Semantics and Pragmatics* 17(5). 1–54.
- Koev, Todor. 2018. Notions of at-issueness. *Language and Linguistics Compass* 12(12). e12306.
- Kratzer, Angelika. 2006. *Decomposing attitude verbs*. Unpublished.
- Kratzer, Angelika. 2016. *Evidential moods in attitude and speech reports*. Unpublished.
- Langendoen, D. Terence & Savin, Harris B. 1971. The projection problem for presuppositions. In Fillmore, Charles J. & Langendoen, D. Terence (eds.), *Studies in linguistic semantics*, 54–60. New York, NY: Irvington.
- Maier, Emar. 2015. Quotation and unquotation in free indirect discourse. *Mind & Language* 30(3). 345–373.
- Maier, Emar. 2017. The pragmatics of attraction. In Saka, Paul & Johnson, Michael (eds.), *The Semantics and Pragmatics of Quotation*, 259–280. Dordrecht: Springer.

- Maier, Emar. 2018. Quotation, demonstration, and attraction in sign language role shift. *Theoretical Linguistics* 44(3-4). 265–276.
- Maier, Emar. 2020. Quotes as complements: A Kratzerian approach. In Bhatt, Rajesh & Frana, Ilaria & Menéndez-Benito, Paula (eds.), *Making worlds accessible: Essays in honor of Angelika Kratzer*, 91–101. Amherst, MA: UMass Amherst.
- Maier, Emar. 2023. Emojis as pictures. *Ergo: An Open Access Journal of Philosophy* 10(11). 310–340.
- Partee, Barbara Hall. 1973. The syntax and semantics of quotation. In Anderson, Stephen R. & Kiparsky, Paul (eds.), *A Festschrift for Moris Hall*, 410–418. New York: Holt, Rinehart and Winston.
- Pearson, Hazel. 2010. A modification of the “Hey, Wait a Minute” test. *Snippets* 22. 7–8.
- Pearson, Hazel. 2012. *The sense of self: Topics in the semantics of de se expressions*. Cambridge, MA: Harvard University. (Doctoral dissertation.)
- Pierini, Francesco. 2021. Emojis and gestures: A new typology. In Grosz, Patrick G. & Martí, Luisa & Pearson, Hazel & Sudo, Yasutada & Zobel, Sarah (eds.), *Proceedings of Sinn und Bedeutung (SuB) 25*, 720–732. London: University College London.
- Plank, Frans. 1986. Über den Personenwechsel und den anderer deiktischer Kategorien in indirekter Rede. *Zeitschrift für germanistische Linguistik* 14(3). 284–308.
- Potts, Christopher. 2005. The logic of conventional implicature. Oxford, UK: Oxford University Press.
- Potts, Christopher. 2007a. The dimensions of quotation. In Barker, Chris & Jacobson, Pauline (eds.), *Direct Compositionality*, 405–431. Oxford: Oxford University Press.
- Potts, Christopher. 2007b. Into the conventional-implicature dimension. *Philosophy Compass* 2(4). 665–679.
- Quine, W. V. 1940. *Mathematical logic*. Cambridge: Harvard University Press.
- Recanati, François. 2001. Open quotation. *Mind* 110(439). 637–687.
- Rett, Jessica. 2021. The semantics of emotive markers and other illocutionary content. *Journal of Semantics* 38. 305–340.
- Saure, Christopher. 2025. *On the interpretation of perspective-dependent expressions in the complement clauses of propositional attitude verbs*. Wuppertal: University of Wuppertal. (Doctoral dissertation.)
- Schlenker, Phillippe. 2003. A plea for monsters. *Theoretical Linguistics* 26(1). 29–120.
- Schlenker, Phillippe. 2004. Context of thought and context of utterance: A note on free indirect discourse and the historical present. *Mind & Language* 19(3). 279–304.
- Schlenker, Phillippe. 2018. Gesture projection and cosuppositions. *Linguistics and Philosophy* 41(3). 295–365.
- Schlenker, Phillippe & Lamberton, Jonathan. 2024. Iconological semantics. *Linguistics and Philosophy* 47. 779–838.
- Shanon, Benny. 1976. On the two kinds of presuppositions in natural language. *Foundations of Language* 14(2). 247–249.
- Sharvit, Yael. 2008. The Puzzle of Free Indirect Discourse. *Linguistics and Philosophy* 31(3). 353–395.
- Simons, Mandy & Tonhauser, Judith & Beaver, David & Roberts, Craige. 2010. What projects and why. In Li, Nan & Lutz, David (eds.), *Proceedings of Semantics and Linguistic Theory (SALT) 20*, 309–327. Ithaca, NY: CLC Publications.
- Stec, Kashmiri. 2012. Meaningful shifts: A review of viewpoint markers in co-speech gesture and sign language. *Gesture* 12(3). 327–360.
- Stec, Kashmiri. 2016. *Visible quotation: The multimodal expression of viewpoint*. Groningen: Rijksuniversiteit Groningen. (Doctoral dissertation.)
- Stokke, Andreas. 2021. Protagonist projection, character focus and mixed quotation. In Maier, Emar & Stokke, Andreas (eds.), *The language of fiction*. 379–404. Oxford: Oxford University Press.

- Stokke, Andreas & Derek Ball. 2025. Quotative be like. *Synthese* 21. 1–26.
- Syrett, Kristen & Koev, Todor. 2015. Experimental evidence for the truth conditional contribution and shifting information status of appositives. *Journal of Semantics* 32(3). 525–577.
- Tonhauser, Judith. 2012. Diagnosing (not-)at-issue content. In Bogal-Allbritten, Elizabeth (ed.), *Proceedings of Semantics of Under-Represented Languages of the Americas (SULA) 6*, 239–254.
- Umbach, Carla & Gust, Helmar. 2014. Similarity demonstratives. *Lingua* 149. 74–93.
- Von Fintel, Kai. 2004. Would you believe it? The king of France is back! In Reimer, Marga & Bezuidenhout, Anne (eds.), *Descriptions and Beyond*. 315–347. Oxford: Oxford University Press.
- Walter, Sebastian. 2025a. Indirect discourse as mixed quotation? An experimental investigation. In Knowlton, Tyler & Schwarz, Florian & Papafragou, Anna (eds.), *Proceedings of Experiments in Linguistic Meaning (ELM) 3*, 423–434. Philadelphia: University of Pennsylvania.
- Walter, Sebastian. 2025b. A mixed-quotational account of indirect discourse: Evidence from self-pointing gestures. In Zhang, Yao & Zhao, Fengyue & Cho, Youngdong & Wu, Yifan (eds.), *Proceedings of Semantics and Linguistic Theory (SALT) 34*, 88–109. Rochester, NY: University of Rochester.
- Walter, Sebastian. 2025c. *Interactions of perspective-taking in gesture and speech*. Frankfurt/Main: Goethe University Frankfurt. (Doctoral dissertation.)
- Walter, Sebastian & Stefan Hinterwimmer. 2025a. An experimental investigation of perspective alignment in gesture and speech. In Knowlton, Tyler & Schwarz, Florian & Papafragou Anna (eds.), *Proceedings of Experiments in Linguistic Meaning (ELM) 3*, 411–423. Philadelphia, PA: University of Pennsylvania.
- Walter, Sebastian & Hinterwimmer, Stefan. 2025b. Shifted face emoji in indirect discourse: A mixed-quotational approach. In Longo, Federica & Panizza, Daniela (eds.), *Proceedings of Sinn und Bedeutung (SuB) 29*, 1662–1680. Noto, Italy: University of Messina.
- Wade, Elizabeth & Clark, Herbert H. 1993. Reproduction and demonstration in quotations. *Journal of Memory and Language* 32(6). 805–819.

Sebastian Walter
Goethe University
Norbert-Wollheim-Platz 1
Postfach 2
60629 Frankfurt am Main
Germany
e-mail: s.walter@em.uni-frankfurt.de

Stefan Hinterwimmer
University of Hamburg
Von-Melle-Park 6
Postfach 15
20148 Hamburg
Germany
e-mail: stefan.hinterwimmer@uni-hamburg.de