Routledge
Taylor & Francis Group

Research on Language and Social Interaction

ISSN: 0835-1813 (Print) 1532-7973 (Online) Journal homepage: www.tandfonline.com/journals/hrls20

Negotiating shared understanding: Coding repair in social interaction

Patrick G.T. Healey, Julian Hough, Dirk vom Lehn, Elif Ecem Özkan & Rose McCabe

To cite this article: Patrick G.T. Healey, Julian Hough, Dirk vom Lehn, Elif Ecem Özkan & Rose McCabe (2025) Negotiating shared understanding: Coding repair in social interaction, Research on Language and Social Interaction, 58:3, 281-302, DOI: 10.1080/08351813.2025.2528495

To link to this article: https://doi.org/10.1080/08351813.2025.2528495

Published online: 02 Sep 2025.

RESEARCH ON LANGUAGE AND SOCIAL INTERACTION
2025, VOL. 58, NO. 3, 281–302
https://doi.org/10.1080/08351813.2025.2528495

Routledge
Taylor & Francis Group

Negotiating shared understanding: Coding repair in social interaction

Patrick G.T. Healey[a], Julian Hough[b], Dirk vom Lehn[a], Elif Ecem Özkan[c], and Rose McCabe[d]

[a]School of Electronic Engineering and Computer Science, Queen Mary, University of London, London, UK; [b]School of Mathematics and Computer Science, Swansea University, Swansea, UK; [c]Kings Business School, King's College London, London, UK; [d]School of Health and Psychological Sciences, City St Georges, University of London, London, UK

ABSTRACT

Repair–the process of detecting and responding to problems with speaking, hearing or understanding in conversation–is the focus of a range of established coding protocols. We discuss the practical process of developing and applying these protocols to a range of verbal and non-verbal repair phenomena. Coding protocols necessarily trade detail for generalization. We consider four payoffs: (i) practical–reduced effort and increased speed and scale of data analysis; ii) empirical–quantitative insights into the distribution of repairs that support comparative analysis and applications such as detecting Alzheimer's Disease; iii) computational –enabling automatic detection of verbal and non-verbal repairs for corpus analysis, dialogue systems development, and to support selective experiments on repair processes; (iv) interdisciplinary–the process of protocol development provides a methodological bridge between qualitative and quantitative disciplines that can foster new insights into how repairs work. Data are in English and German.

Introduction

One of the first general phenomena identified by conversation analysts was *repair*: the ways in which people detect and respond to problems with speaking, hearing or understanding in conversation (Jefferson, 1972; Schegloff et al., 1977). Three features make repair a natural candidate for development of coding systems: it is pervasive in natural conversation (Schegloff et al., 1977); it has a formal organisation considered to be largely independent of the type of problem being dealt with (Jefferson, 1972; Schegloff et al., 1977; Schegloff, 1987); and there is an extensive literature describing its structural and procedural organisation (see Kitzinger, 2012).

Several protocols for coding repair have been published over the last 20 years (e.g., Healey & Thirlwell, 2002; Healey et al., 2005; Stivers & Enfield, 2010; Kendrick, 2015; Dingemanse et al., 2016) and the relative maturity of these protocols makes them a useful test case for exploring the strengths and weaknesses of coding as a method for conversation analysis. We explore a variety of practical, technical and conceptual challenges involved in developing coding protocols. These include which phenomena to include, how coding criteria are defined, what steps are taken to standardize use and how to assess their effectiveness (see also Stivers & Rossi, 2025/this issue).

The development of reliable, generalisable coding criteria inevitably involves compromising details of the target phenomena (c.f. Schegloff, 1993). The simplest justification for these compromises is that they enable quantified estimates of the distribution of different repair phenomena (e.g., Colman & Healey, 2011; Dingemanse et al., 2015; Kendrick, 2015) and direct comparisons of their frequency in different contexts (e.g., Colman & Healey, 2011; Dingemanse et al., 2015). These comparisons can be useful in applications such as predicting diagnoses and treatment outcomes in clinical contexts (e.g., McCabe et al., 2013; Nasreen et al., 2021).

CONTACT Patrick G.T. Healey p.healey@qmul.ac.uk

However, we argue that coding protocols do more than enable quantification and comparison. They also involve the production of forms of formal and procedural description that can, in principle, be made machine readable i.e., can be computationally manipulated and transformed. This makes it possible to build tools that automatically detect and classify repair phenomena (see e.g., Hough & Purver, 2014; Purver et al., 2018); carry out real-time monitoring of problems of understanding in conversation (e.g., Borges et al., 2019); and selectively test the effects of different repair types on live interactions (e.g., Healey et al., 2003; Healey et al., 2018; Mills & Redeker, 2023). This extends the range of tools, empirical methods and types of evidence available to researchers interested in repair.

Coding systems can also provide a methodological bridge between disciplines; they foster a (partial) alignment of empirical practice that can facilitate dialogue between research in conversation analysis, psycholinguistics and formal pragmatics; disciplines that are built on very different foundational assumptions (Healey et. al., 2018). We hope to show that this dialogue is productive, leading to new insights about the character, form, scale and effects of repair in human interaction (see also Hayano, 2025/this issue; Stivers & Rossi, 2025/this issue).

We introduce the basic organisation of repair using four illustrative examples and then show how some basic features of these examples can be captured using a simple coding protocol. We extend these examples to consider more complex forms of repair coding and discuss how researchers have addressed the issues of *validity* and *reliability* of coding. The next section explores some important challenges involved in defining coding criteria, including borderline cases and multimodality. We balance this against some of the payoffs that coding protocols can provide including an overview of recent work on automated repair coding systems and some wider applications of repair coding systems including quantitative analysis of clinical interactions, computational methods for realtime detection of verbal and non-verbal repairs, and experimental tests of the effects of repairs on interaction. The paper ends with a summary of practical recommendations for coding repair and some final conclusions.

The organisation of repair

The CA interest in repair emerges from Sacks' discussions of forms of clarification such as *correction invitation devices* and *appendor questions* and the "special places" where they occur in conversation (*Lectures on Conversation*, Fall 1964-Spring 1965). Sacks' initial interest in clarification arose in the context of analysing membership categorisation. However, by 1967 the distinctive structure of correction sequences had emerged as a distinct analytic interest for a new empirical account of intersubjectivity built on how "misunderstandings are discoverable and discovered, and remedied" (Fall, 1967, p. 735.).

Jefferson (1972) published the first general characterisation of *correction sequences* including: the use of repetition to locate sources of trouble, the importance of distinguishing between who identifies a trouble source and corrects it, the preference for people to correct their own utterances, the use of multi-turn sequences for performing corrections (e.g., *misapprehension sequences*), and the different structures available for how people choose to correct (or not). Sacks, Jefferson and Schegloff (1974) introduce the term *repairs* for this class of phenomena, identifying them as a "grossly apparent" feature of natural conversation.

The most common repair operations used by English speakers involve replacing parts of a turn with alternatives (Schegloff, 2013). This can include deleting, searching, inserting, parenthesizing, aborting, and reordering (Wilkinson & Weatherall, 2011). There are also larger-scale operations, often prompted by a clarification request, in which a whole turn may be revised or reformulated (see Kitzinger, 2012 for a comprehensive overview).

Basic repair operations can be classified on three dimensions:

- **Initiation**: who signals a problem? The person producing the problem turn (*Self*) or a recipient (*Other*)?
- **Response**: who produces the proposed repair (Self or Other)?
- **Position**: where in the sequence of turns do the repair initiation and repair response occur? During production of the turn (*Position 1*), immediately after the production of the turn (*Transition Space*),

in the next relevant turn following the problem (*Position 2*), in the next relevant turn following Position 2 (*Position 3*) or in the next relevant turn following position 3 (*Position 4*).

When a recipient directly prompts someone to clarify a problem turn, this is classified as a *Next Turn Repair Initiation* or *NTRI*.

Example repairs

To illustrate how this classification works, and the operation of a simple repair coding protocol, we introduce four different types of repair, illustrated using multimodal examples drawn from the German subcorpus of the DUEL (DisflUencies, Exclamations and Laughter) corpus (Hough et al., 2016).[1] The subcorpus consists of 10 dyadic conversations, 8-15 minutes long. The participants' task is to design a dream apartment to share, given a fixed footprint and a budget of €500,000. Participants were students of Bielefeld University who consented for data to be presented for research purposes. The corpus includes two camera angles for each participant and separate audio recordings from lapel microphones (see below).

Extract 1: First Position Self-initiated Self-repair (↑ = Rising intonation, (0.3) = pauses in 10ths of a second, [] indicate actions overlapping with the turn above)

```
01    A: Ba↑d wollten wir sechzig machen wären wir bei
         hundert↑zeh:n (0.3) [nee hundertdreißig↑
         Ba↑throom we wanted to allocate sixty, we would be at a
         hundred and ↑te:n (0.3) no a hundred and thirty
                             [nod]
```

Figure 1: Extract 1 video screenshots of A (left of picture) and B (right of picture). Participants are A and B in conversation r1 in DUEL (de). Timestamps: (a) 20:45.5 (b) 20:46.5

In Extract 1, participant A initially suggests they have used up "hundert↑zeh:n (sqm)" of the allocated space. This is followed by a brief pause of 0.3s and then a negation, "nee" accompanied by a forward movement of the head (Figure 1a-b). This suggests a repair initiation, rejecting the initial formulation and then replacing it with "hundertdreißig [sqm]".

[1]Anonymized transcripts and laughter annotations are available from https://github.com/clp-research/DUEL

Extract 2: Second position self-initiated other repair (collaborative completion)[2]

```
01    D:    Natürlich mit 'ner          REPAIR INITIATION LINES 1-4
            Absolutely with a

02          (---[----------)  (1.4)

03              [<><[><>]<>(.)<>     [gesture lasts 3.8s until line 7]

04          (----------)            (1.0)

05    C:            [Mit 'ner Panorama-Tapete          PROVIDES REPAIR
                    [With a landscape wallpaper

06          [auf jeden Fall
            [definitely

07    D:    [richtig                                   CONFIRMS REPAIR
            [right
```

In Extract 2, the participants discuss how to furnish and decorate their flat. Participant D starts to suggest "Natürlich mit 'ner" but then pauses and initiates a gesture with both hands in front of his lap and shifts his gaze from the 'middle distance' to Participant C (line 3, Figure 2a). This prompts C to produce a repair that completes D's utterance, "Mit 'ner Panorama-Tapete" (line 5) by recycling the beginning "Mit 'ner" and producing a gesture in front of her chest that echoes D's gesture (Figure 2b). C indicates alignment with D who produces a smile suggesting he accepts the proposed completion. C emphasises the importance of this feature "auf jeden Fall" (line 6) and D provides an additional confirmation "richtig" (line 7).

Extract 3: Second position NTRI followed by Third Position Other-Initiated Self-repair

```
01    C:    Und [am besten quadratisch ne?
            And preferably square-shaped right?
                [Gesture onset
            (0.8)

02    D:    Der spiegel.
            The mirror
                  C gesture end]

03    C:    Nein (0.2) das zimmer [laughter].
            No (0.2) the room [laughter].
```

In Extract 3 the participants discuss a room in the flat. After an exchange about whether to purchase a wardrobe with an inbuilt mirror or just a mirror, C suggests "preferably square-shaped right?" (line 1) and produces an iconic gesture of a square by opening her arms wide in front of her body (Figure 2c). D asks if "quadratisch" ("square-shaped") refers to "Der Spiegel" ("the mirror", line 2) and C repairs "Nein" (line 3) followed by "das Zimmer" ("the room") and laughter (line 3). Thus, C displays her recognition that D misunderstood what she was referring to. D orients to C's response

[2]Transcription annotation: '-', a tenth of a second; '<>', B's gesture, '[', overlapping conduct.

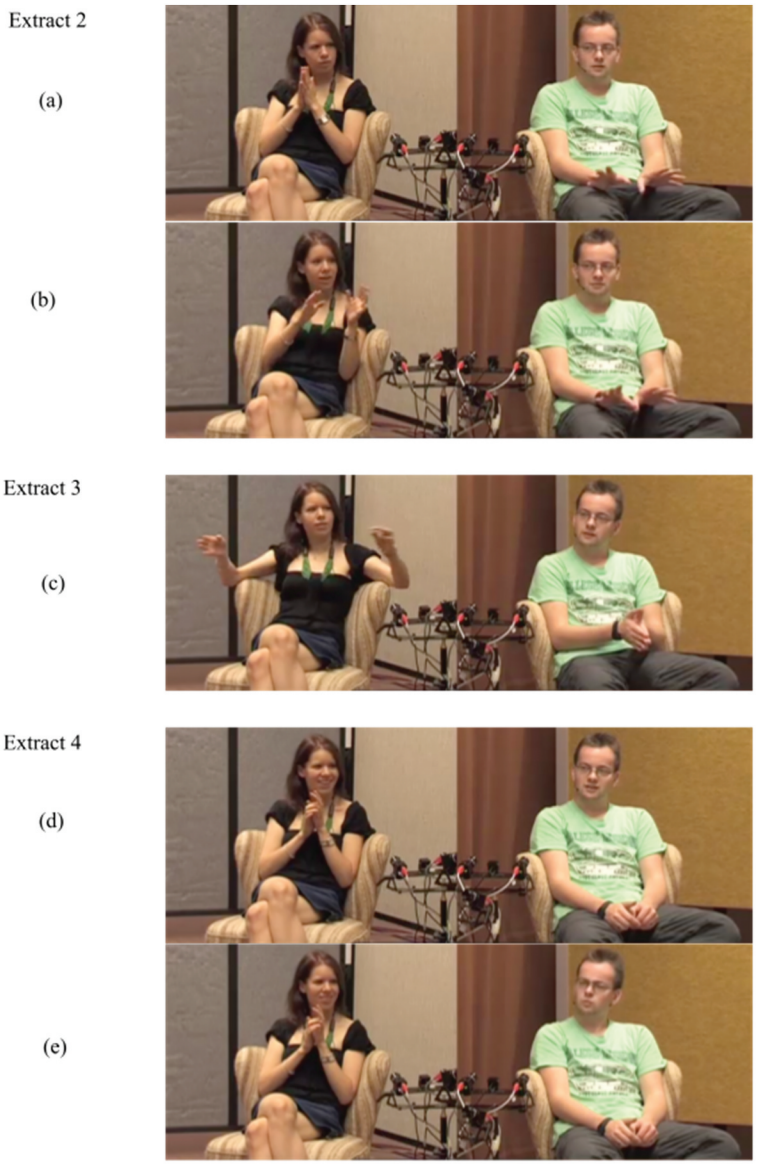

Figure 2: Extracts 2-4 video screenshots of C (left of picture) and D (right of picture). Participants are in conversation r4 in DUEL (de). Timestamps: (a) 13:34.2 (b) 13:35.0 (c) 12:13.9 (d) 15:17.6 (e) 15:18.6

Extract 4: Second position NTRI (non-verbal) followed by Third position Other-initiated Self-Repair (extension)

```
01    C:    und mit einem- mit vielleicht sachen die nicht (hhhh)
            aus (hh) ein (hh) ander brechen
            and maybe with a- with things that don't fall apart

02    D:    [puzzled face lasts until end line 3] NON-VERBAL RI

03    C:    .hh [so wie die bank (hhhhh)?
                [like the bench?
            (.)
            bei uns (hhhhh)
            at ours                                PROVIDES REPAIR

04    D:    Ah ja richtig.
            Oh yeah right.
```

not by joining in her laughter but by questioning C's suggestion that the room should be square-shaped (line 1).

In Extract 4, D produces a puzzled face displaying he does not understand what C is referring to when saying she would like to have furniture that does not fall apart, "vielleicht Sachen die nicht (hhh) aus(hh)ein(hh)ander brechen" (line 1; Figure 2d-e). D's puzzled display prompts C to clarify her reference. Having laughed previously while talking about the items in their current flat falling apart, she produces an audible inbreath and then points out the faulty item in their current flat by saying "wie die Bank" ("like the bench" line 3), further incrementing the clarification with "bei uns" ("at ours"). D then displays alignment, "Ah ja richtig" (line 4), showing that he understands.

Coding repairs with an example protocol

It is useful to reflect on how the four examples above were developed through an iterative process of selection and refinement. The two authors most familiar with the corpus searched the DUEL videos. Ten example repairs were selected and discussed in a series of meetings to assess their quality and clarity. Consensus on the initial analyses was reached through discussion using the transcript and videos as evidence. One example was rejected because of a lack of agreement. The existing DUEL transcriptions were checked by a native German speaker. The final examples were chosen as they combine basic verbal and non-verbal features of repair. This process took approximately twelve hours.

This brief description of our process is, we believe, broadly representative of how qualitative analyses are typically selected and presented. The principal advantages are that it provides a thorough, context-sensitive and detailed characterisation of specific instances of repair. Its principal weakness is that we are likely to be drawn to particularly clear, interesting or controversial examples. As a result, it is unclear if the chosen examples are representative of the phenomena of interest and also unclear if the process of choosing them is reproducible (e.g., by different groups looking at the same extracts). Coding protocols attempt to address these issues by making coding criteria explicit, standardizing the process of applying the codes and testing the validity and reliability of the resulting classification (see below).

To illustrate the process of applying a coding protocol to the examples above, we use a simple coding scheme that was the first attempt to capture the basic structure of the repair space (Healey & Thirwell, 2002; Healey et al., 2005). Figure 3 shows the protocol consists of a binary branching tree made up of a series of yes/no questions focused on locating changes that "edit, amend, or reprise" all or part of an utterance. These questions are applied to each turn in a transcript and once a repair is identified, the line number and category (grey box) is recorded. Each coded repair is treated as 'removed' from the transcript and the protocol is applied iteratively until no more instances are found. Backward and forward-looking questions capture some sequential relationships and reflect the way that a particular contribution is sometimes only classified as, say, a repair initiation because of the type of response it receives; sometimes reversing an initial classification.

Following this process, the protocol classifies Extract 1 as position 1, self-initiated, self-repair (formulation), and Extract 2 as self-initiated position 2 other-repair. Extract 3 is more complex because of the more equivocal nature of D's "der spiegel". If it is understood as proposing a revision of C's reference to "quadratisch" it is directly classified as a position 2 next-turn repair initiator (NTRI) and, subsequently C's "Nein das Zimmer" will be coded as position 3, other-initiated self-repair. On the other hand, if D's "der spiegel" is not understood as a repair initiation but rather, say, as a confirmation of C's "quadratisch" then "Nein das Zimmer" will be classified as a position 3 self-initiated self-repair. Extract 4 is classified as an NTRI because non-verbal signals

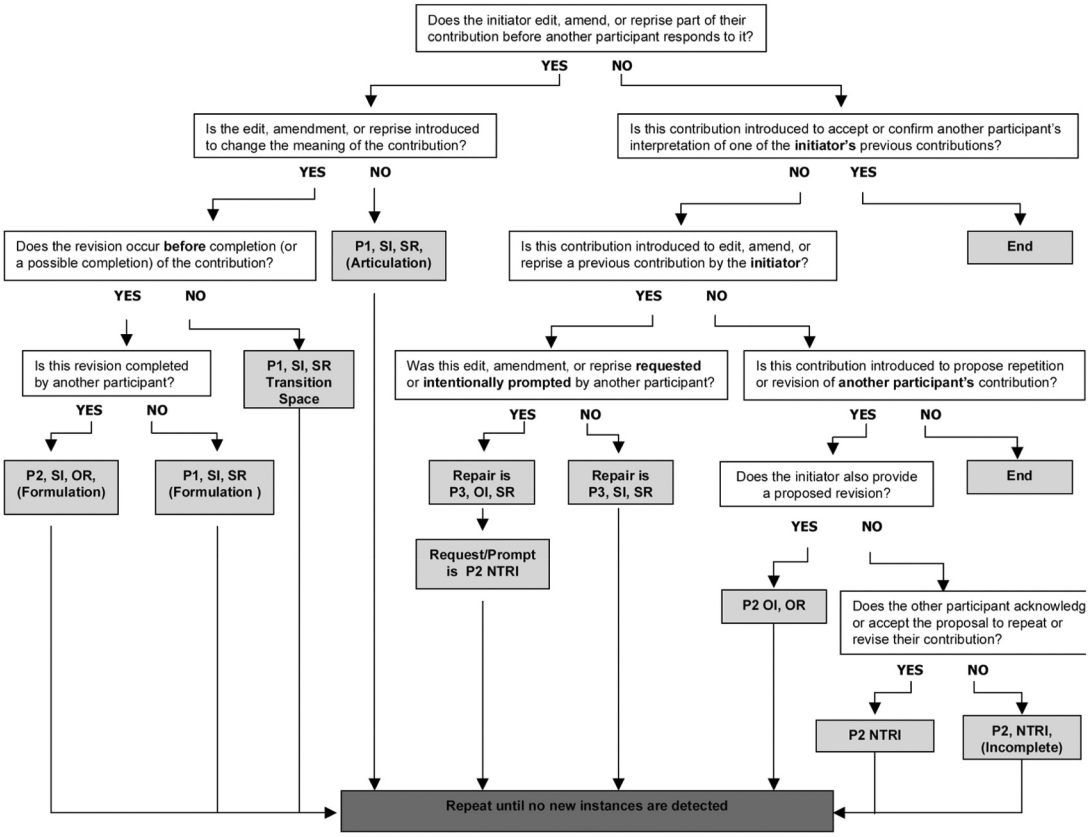

Figure 3: Basic Annotation Protocol from Healey et al. 2005. (S/O = Self/Other; R/I = Initiation/Repair; P1-3 = Positions 1-3, NTRI = Next Turn Repair Initiator).

are included in this protocol as multi-modal contributions and C's "like the bench" is then classified as a position 3, other initiated self-repair.

This example illustrates some potential advantages of using a protocol. It attempts to minimise the need for detailed judgements about context and to maximise the potential for comparison and generalisation. Some familiarity with the structure of conversation and judgements about meaning are still required e.g., whether a repair changes the meaning of (part of) a turn. However, it requires less specific expertise, is less labour intensive and can scale more easily to large datasets. For example, coding the four examples above takes less than an hour, including instruction. It can be applied independently, and the resulting categorisations do not normally require discussion (see reliability and validity below).

The most obvious limitation of the protocol is that it is coarse grained. Around 75% of the classifications of repair it produces align with those of an expert (Healey et al., 2005, see below) but this leaves a substantial number of cases misclassified. The protocol draws extensively on previous qualitative research but glosses over distinctions between different repair operations, including, for example, different kinds of clarification questions (cf. Purver, 2002; 2004). It also does not attempt to code what kind of action a repair might be designed to achieve (cf. Dingemanse et al., 2016; Deppermann & Gubina, 2025/this issue). The protocol also intentionally ignores some phenomena that, arguably, are too rare to be useful for quantitative analysis e.g., position 4 repairs. These limitations illustrate the trade-off between the aims of standardized categorisation and sensitivity to context.

Protocols for coding repair

We now describe a range of protocols that have attempted to capture various different aspects of repair. The concept of repair is broad. It has progressively widened from correction of problems with references to include any aspects of a turn that are treated as problematic by participants (Schegloff et al., 1977; Schegloff, 1987). Table 1 summarizes a variety of different repair coding schemes that address different aspects of the phenomena. No single coding protocol that we are aware of attempts to capture all repair phenomena. For example, the protocol above (Figure 3: Healey et al., 2005) was designed to capture the general shape of the repair space but not the details (cf. Schegloff, 1992). Hough et al. (2015) focus on self-repairs while Purver et al. (2003, 2004) and Dingemanse et al. (2016) focus on clarification questions or other-initiated repairs.

Table 1 includes protocols from conversation analysis, psycholinguistics, formal pragmatics and computational linguistics. Although focused on similar phenomena, this work has often developed in parallel with only limited contacts between the disciplines. Consequently, basic concepts, scope and terminology vary. We summarise some of the main commonalities and contrasts below.

Trouble sources

All schemes identify a fragment, word, phrase, turn or non-verbal contribution that is the object of a repair, but multiple terms are used: *product items* (Jefferson, 1972), *repairable* (e.g., Sacks et al., 1974), *reparandum* (e.g., Shriberg, 1994), *source* (e.g., Purver, 2004), *antecedent* (e.g., Rodriguez & Schlangen, 2004), *trouble-source turn/T-1* (e.g., Kendrick, 2015).

Repair markers or editing phrases

Many protocols use specific *markers* (Schegloff, 1992) that can signal a repair in progress, e.g. "I mean", "umm" and "err". In schemes influenced by psycholinguistics, these are classified as *editing phrases* and/or *filled pauses* (e.g., Ginzburg et al., 2014) and in computational linguistics as *interregna* (e.g., Shriberg, 1994).

Repair solution

The changes ultimately made to a trouble source and accepted during repair have been variously termed *correction* (e.g., Jefferson, 1972) *repair* (e.g., Sacks et al., 1974; Shriberg, 1994), *repair-outcome* (e.g., Schegloff et al., 1977), *alteration* (e.g., Ginzburg et al., 2014) and *repair solution* (e.g., Dingemanse et al., 2016).

Specificity and repairable unit

Repair initiations vary in how precisely they identify or locate a trouble source. In conversation analysis this was introduced as the concept of *specificity* or the power to "locate a repairable" (Schegloff et al., 1977, p. 369). For example, a "Huh?" is low specificity because it might signal a wide range of potential problems whereas a "Who?" or "When?" signal specific problems with references. In its original formulation specificity combined structural criteria based on the form of the trouble source e.g., initiations using partial or full repetition, and semantic criteria, e.g., initiations using wh questions or paraphrase. Rodriguez and Schlangen (2004) adopt a structural notion and code the *extent* of clarification (initiation) as a binary feature identifying either a constituent or a whole turn. Some recent protocols use a more general semantic distinction between *restricted* (specific) and *unrestricted* (non-specific) clarifications to simplify coding across languages (e.g., Kendrick, 2015; Dingemanse et al., 2016). Examples are provided in Table 2.

Table 1: Overview of Repair Coding Protocols (OIR = Other Initiated Repair, SR = Self-Repair).

Scheme	Scope	Validity	Reliability	Procedure	Multimodal	Sequence Based
Bortfield et al. (2001)	Lexical repetition, syntactic restarts, fillers and editing expressions	Analytic criteria, not directly tested	Average across categories excluding editing expressions: Cohen's K = 0.91	Annotation scheme with criteria and examples	N	N
Healey and Thirlwell (2002)/ Healey et al. (2005)	Whole repair space	75% agreement with documented examples from the literature	Average across categories: Cohen's K = 0.76	Binary branching tree protocol	Y	Y - backward looking criteria
Purver (2004)/Purver et al. (2003)	OIR: Clarification requests	Tested on conversations from the British National Corpus (BNC)	Average for Form: Cohen's K = 0.78. Average for Reading /Interpretation = 0.75	Binary Branching Tree Protocol	N	Y - backward looking criteria.
Rodriguez and Schlangen (2004)	OIR: Clarification requests in German	Tested on a corpus of German task-oriented dialogues	Complete annotation by one author. "source of the problem" annotated by a second annotator: Cohen's K = 0.70	Annotation scheme with criteria and examples	N - but includes intonation feature (rising and falling)	Y - backward looking with "relation to the antecedent" feature
Hough et al. (2015) / Shriberg (1994)	First position SR	Tested per-word on transcripts with 3 annotators	Reparandum words, = 95% agreement (K-free =0.90); Repair words 97% agreement (K-free=0.94)	Annotation scheme with written criteria and examples	N	Y - on word-level with backwards looking criteria
Kendrick (2015)/ Dingemanse, Kendrick and Enfield (2016)	OIR	Analytic criteria. Not directly tested	Not reported	Step-wise procedure plus written criteria	Y	Y - Positions: T-1, T0, T +1
Manrique (2016)	OIR: Visual signals (head, face, posture) LSA	Analytic criteria. Not directly tested	Not reported.	Transcription scheme specific to non-manual markers Written criteria and examples	N - Visual signals only but various body parts	Y - Positions: T-1, T0, T +1
Howes and Eshghi (2021)	OIR: Clarification requests	Tested on 9 conversations from the BNC with 2 annotators	Cohen's K =0.64–0.93 (average 0.78) for all classes inc. CRs,	Binary branching tree protocol	N	Y - Backward looking criteria
Homke et al. (2025)	OIR: Eyebrow raise vs. Eyebrow Furrow	Tested on 59 random minutes with 2 annotators	Cohen's K = 0.88	Annotation scheme with criteria and examples	N - visual signals only	Y - Positions: T-1, T0, T +1

The DUEL scheme for first position self-repairs (Hough et al., 2015, based on Shriberg, 1994) codes trouble sources at the word level. It separates the *reparandum* up to the +, an optional *interregnum* in {} brackets with an *F* for filled pauses, and a *repair* (the words after the *interruption point* + up to the closing bracket) as illustrated by the following example: "John (likes + {F uh} loves) Mary" where "likes" is the reparandum, "uh" is the interregnum and "loves" is the repair.

Severity

Some schemes code how much change a repair solution makes to an utterance. Rodriguez and Schlangen (2004) distinguish the degree of repetition of previous material from a broader confirmation of an alternative interpretation or *hypothesis*. Healey et al. (2005) suggest measuring the extent of repairs, e.g., how much surface material -words, drawings or gestures- are edited or replaced, as an index of severity.

Form

The most useful features of repair for the purposes of coding are recurrent, recognisable *forms* that can be identified independently of specific contexts or the type of trouble (Jefferson, 1972, p. 321; Sacks et al. 1974, p. 717–718). However, the way *form* is characterized varies significantly between disciplines.

CA emphasises the form of the turn sequences or *positions* that make up the local *repair space* (Schegloff, 1992) and highlights differences in the form of turns within that space. For example, *third position* repairs are located at least one turn from the trouble source and typically employ a distinctive format (although see Extract 3) to introduce a repair including pauses, turn-initial markers ("well"/"oh"), agreement-acceptance particles ("yes, no") and explicit edit terms ("I mean"). Dingemanse et al. (2016) classify the repair space for other initiated repairs using *T0* for the repair initiation, *T-1* for the trouble source, *T+1* for the repair solution and *T+2* for accepting the repair (see e.g., Extract 2, line 7). Strictly, both *position* and *Tn* are semantic or pragmatic not formal concepts because they are defined in terms of the (analyst's judgement of) *relevance* of turns to a repair.

In contrast to this formal pragmatics and psycho/computational linguistics emphasise syntactic and lexical categories both at the word level (e.g., Shriberg, 1994; Hough et al., 2015: see (1) above) and turn level including structural relations, such as ellipsis, within and across turns. Clarification questions are classified into formal types such as *polar question, wh-question/sluice* and analysis focuses on the extent to which these syntactic and lexical categories match the form of the trouble source (Purver et al., 2003; Rodriguez & Schlangen, 2004).[3] Table 2 summarises some interrelationships between CA and formal pragmatic analyses of *other-initiations* e.g., *next-turn repair initiator (NTRI)* or *clarification requests*.

Other useful formal cues are intonation e.g., *boundary tone* (rising vs. falling, Rodriguez & Schlangen, 2004), *unfilled pauses* (mid-unit silences marked with .), partial words, non-standard forms of words and laughter (Hough et al., 2015).

Standardization

The claims of a coding protocol to be systematic and standardized depend on their validity and reliability.

[3]It is interesting to note that Sacks initially considered repairs as analogous to Chomskian syntactic transformations but defined on turns rather than sentences (lecture 9, p. 138, Winter 1969).

Table 2: Comparison of Types of Clarification Questions and Other-Initiated Repairs.

Formal Pragmatics Classification	Format	Conversation Analysis (CA) Classification.	Illustrative Examples (adapted from Purver et al. 2003)
Non-reprise Clarification	Full question	*Other-Initiated NTRI Second Position. Open. No repeat.*	"What do you mean?"/ "What did you say?"
Conventional	Single word clarification	*Open Class Repair Initiator. Unrestricted. No repeat.*	"Eh?" "Sorry?" "What?" "Pardon?"
Reprise Sentence/Reprise Interrogative	Sentential repeat of (part of) prior turn. Not necessarily verbatim.	*Other-Initiated NTRI, Second Position. Restricted. Partial Repeat.*	A: "I spoke to him on Wednesday, I phoned him". G: **"You phoned him?"** O: "Phoned him"
Reprise Fragment	Verbatim repeat of non-sentential part of prior utterance.	*Other-Initiated NTRI, Second Position. Restricted. Partial Repeat.*	A: "There's only two people in the class" B: **"Two people?"** A: "For cookery, yeah.".
Reprise Sluice	Wh Question.	*Other-Initiated, NTRI, Second Position./Category Specific Interrogatives. Restricted. No Repeat.*	"Who?" "What?" "When?" "Where?"
Gap	Repetition of a word preceding a trouble source	*Other-Initiated NTRI, Second Position. Restricted. Partial Repeat.*	A: Can I have some toast please? B: **Some?** A: Toast
Gap Filler	Possible completion of a prior utterance.	*Self-Initiated Other Repair, Second Position.*	A: "if, if you try and do enchiladas or" B: "Mhm" A: "erm" B: **"Tacos?"** A: "tacos".

Validity

A fundamental question for any empirical analysis is validity: how faithfully does it capture the "underlying natural phenomena" (Schegloff, 1999, p. 468)? Coding protocols approach this issue in three broad ways.

Firstly, many coding schemes do not assess validity directly (see Table 1). Instead, they argue for the analytic validity of the coding criteria they use. The operation of these criteria is demonstrated for particular cases, but how well they generalise is an open question.

A second, procedural, response is to maximise context sensitivity by starting from a relatively rich understanding of individual cases as part of applying the coding criteria (see e.g., the discussions in Schegloff, 1996; Dingemanse et al., 2016; Clayman & Heritage, 2025/this issue). This process is an essential step in protocol development but if a bespoke approach is used for each subsequent analysis this works against the practical advantages that a protocol can bring, i.e., reduced labour and increased speed and scale. It also raises the question of what specific judgements about the context the analysts are making and whether these judgements could be explicitly integrated in the coding instructions.

A third approach, common in psycholinguistics, is to provide a quantitative estimate of validity by assessing a coding protocol against some 'gold-standard' classification. For example, Healey and Thirlwell (2005, see Figure 3) collected a corpus of 76 pre-existing examples of repair analyses from the CA literature and applied their protocol to an unlabelled transcript of each example. The coding procedure assigned 75% of the repairs to the same category as the published analysis. Depending on purpose, this may be sufficient for relatively coarse-grained quantitative comparisons but will not classify all instances correctly.

Quantitative estimates of validity depend on the quality of the 'gold standard' used. Richly detailed qualitative analyses are the best option. However, these analyses are also, to some extent, a moving target. The agreed analysis of published examples can change over time and so do the conventions for analysing repair. Also, as noted above, published examples may be biased towards especially clear, interesting or even controversial examples of repairs. Testing against large collections of examples helps to mitigate this.

Reliability and inter-annotator agreement

A second key question for a standardized protocol is whether different people reliably produce the same classification of the same example. This is essentially a measure of whether the analyses can be independently reproduced. It is also a useful practical way to help 'debug' protocol instructions and coding criteria by identifying which categories are applied consistently and which cause confusion (see e.g., Healey et al., 2005; Howes et al., 2014; König & Pfeiffer, 2025/this issue).

In psycholinguistic research, reliability or inter-annotator agreement (IAA) is normally assessed with Cohen's Kappa.[4] Unlike % agreement, Cohen's Kappa takes agreement by chance into account by calculating how often coders agree relative to how often they would be expected to agree if codes were applied randomly. This is important if, as in the case of repair, the baseline frequency of the different target phenomena varies widely.

Cohen's Kappa is limited, amongst other things, to mutually exclusive categories and to pairs of coders (Carletta, 1996). A variety of other techniques are available. Fleiss's Kappa can be used for multiple coders, and Kappa-free is a measure of agreement which does not assume an underlying distribution of categories (Randolf, 2005). Table 1 summarises the inter-annotator reliability estimates for different protocols where reported.

[4]Cohen's Kappa results are conventionally interpreted as follows (Landis & Koch, 1977):
0.01 – 0.20 slight agreement
0.21 – 0.40 fair agreement
0.41 – 0.60 moderate agreement
0.61 – 0.80 substantial agreement
0.81 – 1.00 almost perfect or perfect agreement
Kappa is in the range [0,1]. A value of 1 implies perfect agreement and values less than 1 imply less than perfect agreement, with agreement at chance levels being 0.

Challenges for analysis

The process of constructing a coding protocol and testing its validity and reliability highlights some of the basic methodological and analytic challenges that repair poses.

Ambiguity of form

Repetition of form, especially repetition of material from the trouble source, is useful for coding because it is relatively easy for humans to identify. Repetition is also especially useful for the development of automated tools to search for repair sequences (e.g., Purver, 2002; Hough & Purver, 2014; Purver et al., 2018). However, no two instances of speech signal or body movement are ever identical, especially when produced by different people. In practice, what counts as a repetition is always relative to the type of representation and transcription used and often judgements about meaning and intent (see also Clayman & Heritage, 2025/this issue). Transcriptions are also often revised in successive analyses. Ultimately, even the 'raw' data such as an audio recording, video, or motion capture– are themselves partially incomplete representations of an interaction.

Ambiguity of function

The function of an utterance cannot be read directly from its form, especially across languages (Dingemanse & Enfield, 2015). This means it is difficult to reliably classify what a repair is doing. There are a wide variety of answers ranging from editing a syllable through correcting an action and even to language change (Healey, 2008).

A variety of *repair operations* have been identified but there is no clear consensus in the literature. CA inventories tend to be the most comprehensive and include deleting, searching, parenthesising, aborting, sequence jumping, recycling, reformatting and reordering (Kitzinger, 2012; Schegloff, 2013). More restricted but overlapping sets of distinctions are found in many coding schemes e.g., *reprise, edit, amend* (Healey et al., 2005) *repetition, substitution* and *deletion* (e.g., Hough, 2015), *repetition, addition, reformulation* and *independent* (Rodriguez & Schlangen, 2004).

There is also a more fundamental difference of emphasis on what these different functions operate on. CA-oriented approaches such as Dingemanse et al. (2016) focus on social action types (i.e., *no other action, surprise/disbelief, disaligning action, non-serious action, other*). Approaches influenced by formal pragmatics focus primarily on the semantic (meaning) updates (e.g., Ginzburg, 2012; Ginzburg et al., 2014). In principle, both functions can be achieved through the same repair operations.

Drew (1997) shows that some *open* (or *reprise sluice*) repair initiations e.g. "what" or "sorry" can be either a simple request for repetition because someone didn't hear or a challenge the appropriateness of a prior utterance (see Kendrick, 2015). Some work invokes a higher functional level of *joint project* (Clark, 1996) that sometimes seems to be addressed by repairs. Schlöder and Fernández (2014, 2015) illustrate this with the following examples (using the example numbers in their article):

(1) Agent: You need a visa
Cust: I do need one?
Agent: Yes you do.
(2) K: for me that is in fact below this
I: why below?
K: yes, it belongs there, all okay

In (1) and (2) there is no obvious misunderstanding of lexical content or meaning and they do not appear to be challenging the appropriateness of the prior statement but rather imply a problem with understanding what the broader project is. Overall, judgements about function are inherently more complex, and context sensitive than other aspects of repair and consequently harder to incorporate

into coding protocols (see also Clayman & Heritage, 2025/this issue; Deppermann & Gubina, 2025/this issue).

Multimodal repairs

Non-verbal signals have an integral role in repair. Gestures and facial expressions supplement verbal signals and sometimes entirely replace them e.g., the puzzled expression that acts as an other-initiation in Extract 4 (see also Kendrick, 2015). Examples of other-initiated repairs using gestures (Andrews 2014; Healey et al., 2015; Floyd et al., 2016; Mortensen, 2016), head tilts (Seo & Koshik, 2010), leaning forward (Rasmussen, 2014), and puzzled expressions (Borges et al., 2019) have all been documented.

Multimodal repairs pose challenges for transcription and coding (Mondada, 2016, 2018). Capturing the detail of non-verbal dynamics requires a multi-layered timeline, sometimes with millisecond resolution, and new transcription tools (e.g., using ELAN Wittenburg et al., 2006, see Figure 4). Manrique and Enfield (2015) and Manrique (2016) present a visual transcription system for head (up/down), facial (ET, eyebrows together) and hand movements. Five horizontal lines represent how facial and manual gestures and question-words combine into composite other-initiations of repair. Suspension of movement or the "hold" phenomenon (Manrique, 2016; Floyd et al., 2016) can also be significant. For example, in extract 4, Participant B's puzzled face repair initiator is held static until Participant A's utterance "bei uns" resolves the problem.

This multi-layered, fine-grained level of detail is necessary partly because non-verbal signals complicate sequence organisation. They can occur in parallel with each other and with verbal contributions and may be produced concurrently by multiple parties. For example, addressees produce precisely timed signals of generic and content-specific feedback, which have concurrent effects on speaker's turn construction (Bavelas et al., 2000). As a result, something that e.g., appears to be a verbal self-initiated, self-repair may turn out to be a non-verbal other-initiated repair.

Analysis of non-verbal signals also creates pressure for spatial forms of data capture and representation (cf. Mondada, 2016). Video is two dimensional, but body movements are three dimensional and, for example, the location of a gesture can be important for its interpretation. Motion capture technologies can directly capture the three-dimensional organisation of body movements and gestures in repairs but have yet to see much use (Healey et al., 2013, 2015). Rasenberg et al. (2022) combine synchronous multi-modal data (audio, video, and motion tracking recordings) to analyse other-initiated repair sequences.

Borderline cases

Perhaps the most fundamental challenge in coding is deciding whether something is a repair at all. Repair is not only about the correction of objectively verifiable misunderstandings or 'errors', it also involves matters that the participants treat as having been problematic even if there is no apparent misunderstanding or 'error'; people can set out "to 'correct' talk that is apparently unblemished" (Schegloff, 2007b: 100). Conversely, apparent problems are sometimes allowed to pass without repair. As a result, the judgment about what people are addressing and whether they are treating it as a repair can be subtle.

It is sometimes unclear whether a repair is prompted by a trouble with understanding what has been said or whether it is just a request for more information or an account. If there are overt features that show the participants themselves treat something as a misunderstanding this is easy to resolve, but sometimes this depends on more implicit judgements (see also Hayano, 2025/this issue). One example is Jefferson's (1987) observation of embedded corrections where words are replaced without any overt signal that there was a problem or that they have been repaired. These cases are not captured by coding criteria based on surface form and, by definition, participants do not orient to them as repairs even though they do incorporate the replaced words into their subsequent language use.

Another recognized borderline case is question formatted news receipts (see Kendrick, 2015; Dingemanse et al., 2016). In these examples, people produce utterances that look like repair initiations (e.g., "you did?"; "really?"; "are you serious?"), which seem to function as expressions of surprise or disbelief rather than a problem with hearing or understanding. These are included in some studies of repair (e.g., Schegloff, 2000, p. 223; Wilkinson & Kitzinger, 2006, p. 169; Schegloff, 2007, p. 155) but excluded in others (Kendrick, 2015; Dingemanse et al., 2016).

Payoffs: Why use a coding protocol?

The development of standardized coding systems always compromises the phenomena of interest to some degree. The test proposed by Schegloff (1993) is whether the "payoffs" resulting from this trade-off are sufficient. What does a standardized coding protocol offer relative to collection building in CA and is worth it for CA researchers? We believe it is and that the benefits go considerably beyond obtaining a quantitative overview.

The simplest payoffs are estimates of the quantitative distribution of different repair types (Bortfield et. al., 2001; Colman & Healey, 2011; Kendrick, 2015). Quantitative estimates provide a more precise articulation of qualitative observations that, for example, repairs are "grossly apparent" in natural conversation (Sacks et al., 1974, p. 700). In large corpora, a verbal repair occurs in approximately one third of turns or once every 36 words and this is likely to be a conservative estimate because of the transcriptions used (Colman & Healey, 2011). Quantitative estimates also allow for direct comparisons of relative proportions of repair types e.g., providing quantitative support for the qualitative ranking of self-initiation as more common than other-initiation and self-repair as more common than other-repair (Schegloff et al., 1977).

A second, more interesting, payoff is that quantitative estimates enable direct comparisons of the distribution of repairs across different contexts. For example, Bortfield et. al. (2001) find self-repairs don't vary much with demographic factors of familiarity or age but do vary with topic and dialogue role. Similarly, Colman and Healey (2011) find significant differences in the frequency and distribution of repair types between natural dialogue and task-oriented dialogue (see also Hough et al., 2015). These findings significantly extend the structural and preference-based explanation of repair distributions (Schegloff et al., 1977), providing evidence for a link between repair type, content and task demands (see also Purver, 2004).

A third payoff is that quantification facilitates comparison across languages (e.g., Hough et al., 2015; König & Pfeiffer, 2025/this issue). For example, Dingemanse et al. (2015) show that specificity and severity of repairs are consistently correlated in a wide variety of languages providing evidence for the operation of a universal interactional infrastructure.

Automating repair detection

A longer-term payoff of developing protocols is that they provide the foundations for automatic, rather than manual, annotation of repair, including verbal and non-verbal components (see Figures 4, 5 and 6). This has the potential to scale to much larger datasets at significantly lower cost. In some cases, it can also provide a finer level of detail than is feasible manually.

Automated repair systems

Two example systems for automatic self-repair detection are `STIR` (Hough & Purver, 2014) and `deep_disfluency` (Hough & Schlangen, 2017). They use incremental, word-by-word processing and output word start and end times and tags for fluent words (f), edit terms (e) or repair starts (rpS) illustrated in Figure 5. These systems use machine learning algorithms trained on annotated corpora and rely on timing information from tools such as Praat (Boersma, 2007) for audio and ELAN for video. Praat does not provide syllable and phoneme timings, but these can be estimated using 'forced alignment' (e.g., Schiel, 1999).

These systems have generally good validity (agreement levels with human annotations). For example, `STIR` identifies self-repairs in transcripts with a per-word and per-utterance accuracy of 85% F1 Score, which approximates Cohen's Kappa (Purver et al., 2018). This falls to 62% (F1) for data unlike its training data (Howes et al., 2014). `deep_disfluency` is a more complex system that can process automatic speech recognition (ASR) results 'live'. This capability is important for any applications that need to detect and respond to repairs in real time. ASR introduces errors and `deep_disfluency` achieves tagging accuracies of 56% (F1) on live speech compared with 80% (F1) on human-transcribed data. Rohanian and Hough (2021) provide an alternative deep learning system which improves to 61% (F1) on ASR. Recent advances in the quality of automatic transcription, including more comprehensive capture of key surface phenomena such as repetitions and filled pauses (see e.g. Bain et al. 2023), will substantially improve the accuracy of automatic repair detection.

An important issue for automated classification systems is the frequency of the target phenomena. They perform well on self-repairs because they are both frequent and less variable in form. Automatic detection of NTRIs and subsequent repair trajectories is more challenging because they are relatively infrequent and their form is more variable (see above). Other-repair detection methods have reached 55% F1 accuracy for clarification requests on transcripts of unscripted dialogues (Purver et al., 2018).

To illustrate how transcripts can be built up using these techniques consider the CA transcription and repair and gesture analysis from Extract 4 (repeated from above):

Extract 4 (repeated)

```
01   C:   und mit einem- mit vielleicht sachen die nicht (hhhh)
          aus(hh)ein(hh)ander brechen
          and maybe with a- with things that don't fall apart

02   D:   [puzzled face lasts until end line 3]    NON-VERBAL RI

03   C:   .hh [so wie die bank (hhhhh)?
              [like the bench ?
          (.)
          bei uns (hhhhh)                          PROVIDES REPAIR
          at ours

04   D:   Ah ja richtig.
          Oh yeah right.
```

Figure 4 shows how the repair and gesture annotation for Extract 4 can be aligned with utterance transcriptions on the real timeline using tools such as ELAN.

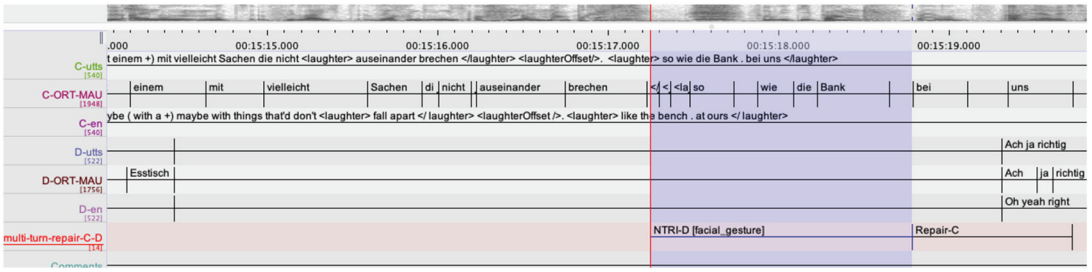

Figure 4: Screenshot of ELAN tiers corresponding to the repair sequence in Extract 4.

Annotations on the real timeline can be combined with automated speech recognition for utterance segmentation and self-repair detection from the speech signal using `STIR` and `deep_disfluency` (see above) and can be trained with Praat TextGrid-style interval annotation (Figure 5)

Onset	15.13.83	15.13.96	15.14.20	15.14.64	15.14.98
Offset	15.13.96	15.14.20	15.14.64	15.14.98	15.15.90
Recognised Token	und	mit	einem	mit	veilleicht
		STIR/ deep_Disfluency ↓			
Repair Type	f	f	f	rpS	f

Figure 5: Pratt TextGrid-style annotation with word timings and automatic repair classification: f = fluent, rpS = repair start.

Automatically generated tiers for hand movements, posture and facial expressions can be added using video analysis tools (e.g. the skeletons and facial movements fitted by MediaPipe Solutions as in Figure 6). These are also aligned to the real time line.

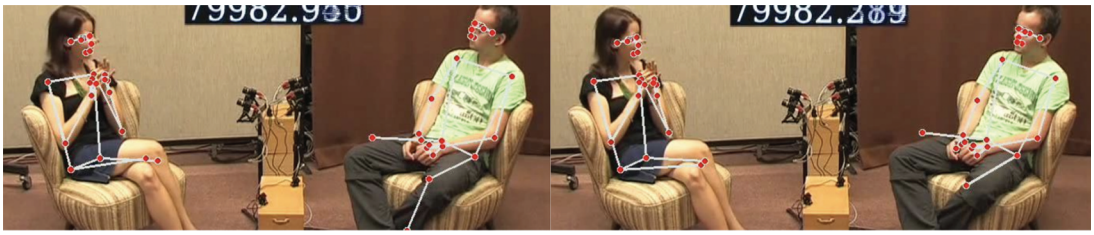

Figure 6: Automatic capture of body movements from Extract 4 using MediaPipe Solutions.

Automatic analysis of multi-modal data for studying repair

The payoffs of automated methods for multimodal data are significantly greater because hand coding is particularly time consuming (e.g., Rasenberg et al., 2022; Dingemanse et al., 2016). Our understanding of what motion features are most effective for discriminating non-verbal forms of repair is in its infancy. However, motion capture from depth cameras has shown that simple measures of hand and head movement, such as height and velocity, are easy to calculate from this data and are systematically related to verbal repairs (e.g., Healey et al., 2013; Healey et al., 2015; Rasenberg et al., 2022).

Recent advances in computer vision tools provide more opportunities for using video data. For example, the landmark model in MediaPipe Pose solutions can detect 33 body key points (Bazarevsky et al., 2020). There is also potential for different modalities to be combined. Ozkan et al. (2021, 2022, 2023) use `deep_disfluency` to identify self-repairs in the DUEL corpus and then MediaPipe for key body point, from both speaker and listener, around each self-repair point frame-by-frame for analysis (see Figure 6).

These automated methods have yielded some early findings on multimodal aspects of repair such as: (i) speech and gesture production tend to ascend and decline in unison across repair types and sequential positions (Rasenberg et al., 2022), (ii) self-repairs correlate with an increase in head and hand height and velocities (Gurion et al., 2020; Ozkan et al., 2021, 2023) and in distances between head and hands of interlocutors (Ozkan et al., 2022).

To date only a handful of studies of automated, multimodal approaches to detecting repairs have been published. However, machine learning techniques are constantly improving the speed, accuracy and resolution of motion data that can be obtained from camera feeds and new sources of non-verbal

data are becoming available from immersive interactions in social virtual realities. An interesting, unintended payoff of trying to develop automated ways of detecting some non-verbal phenomena is that it highlights how even apparently simple movements, such as nods, are highly variable in their physical form (e.g., see Gurion et al., 2020).

Applications of repair coding

Developing these tools has helped to expand interest in repair into fields where quantitative measures are more common. One example is clinical communication where there is interest in the diagnostic properties of repair. For example, rates of repair appear to be condition and symptom specific (e.g., Lake et al., 2011; Leudar et al., 1992). The `deep_disfluency` tagger (see above) finds distinctive repair distributions for people with/without Alzheimer's Disease (AD) (Nasreen et al., 2021; Rohanian, 2023) and its output can be used to improve automatic AD detection from patient speech (Rohanian et al., 2021).

Quantitative estimates also suggest an important role for repair in the quality of doctor-patient communication. In psychiatric consultations more psychiatrist self-repair is associated with a better therapeutic relationship (Themistocleous et al., 2009). McCabe et al. (2013) found that clarification sequences aimed at correcting or understanding what the psychiatrist was saying were associated with better treatment adherence. This suggests that repair - as a mechanism for negotiating shared understanding - is associated with both quality of the relationship and outcome. Psychiatrists who were trained in more effective communication also used 44% more self-repair with their patients (McCabe et al., 2016), suggesting that self-repair may be a useful quantitative index of how hard someone is working at maintaining shared understanding.

Finally, the real-time computational processing of repairs makes it possible to detect and respond to them during live interactions. This has enabled experimental studies that detect and then selectively manipulate repairs e.g. filter them out or substitute them. This approach has provided evidence for the causal effects of repairs in shaping how people adapt to different interlocutors and their role in the emergence of new conventions in real and artificial languages (Healey et al., 2007; Healey, 2008; Healey et al., 2018; Mills & Redeker, 2023). In addition, immersive virtual interactions make it possible to detect and manipulate non-verbal repairs, such as confused or puzzled looks, to test their effects on the trajectory of an interaction (e.g., Borges et al., 2019; Homke et al., 2025). In turn, these experimental studies can feed into the development of dialogue systems and artificial agents that are able to detect and respond to misunderstandings and can use the processing of clarification requests to help a dialogue system to learn words and adapt to different individuals (e.g., Purver, 2002).

Practical recommendations for repair coding

The experience of developing protocols for coding repairs suggests a number of practical recommendations that may be useful for CA practitioners and other researchers interested in repair.

(1) Comb the literature for potentially useful coding criteria i.e., ones that command wide agreement and that identify formal or sequential features that do not rely on subtle judgements about context.
(2) No coding scheme is complete or definitive so be clear about the purposes. What questions are being asked, what kind of evidence is needed to answer them and how much noise can be tolerated in the answers?
(3) Transcription, segmentation and coding decisions interact. If repair coding is done downstream and independently of transcription and segmentation, some phenomena will be missed. This may or may not matter depending on purposes. For languages such as Chinese where orthographic decisions for word segmentation can affect possible repair points, this is especially important.
(4) Annotations of repair phases should be as fine-grained as possible. This could be at the word or phoneme level or even using millisecond or frame-by-frame time-stamps. Tools like Praat or

ELAN (which allow interval-based annotation) can be useful for this and facilitate integration of data from audio, video and motion capture streams.

(5) Relevant nonverbal behaviours, by all parties, should be coded in separate annotation tiers. This allows timing of behaviours in relation to each other and to speech to be extracted more easily, e.g., as in the non-verbal repair initiator in Extract 4.

(6) Testing of a repair coding scheme, ideally using independent annotation and inter-annotator agreement on a 'gold standard', provides the strongest evidence for reproducibility of results. This also applies to automatic coding methods which need to be validated against human annotations.

(7) Strictly exclusive coding categories can be problematic for agreement measures and for ambiguous cases. Allowing multi-level coding so multiple repair events can be tagged over overlapping intervals is important (see Figures 4, 5 and 6). This is possible using binary coding schemes indicating the presence of one or more repair types (like Healey et al., 2005), allowing embedded repair annotation (Hough et al., 2015) and appropriate annotation tools.

Conclusions

Manual and automatic coding protocols are expanding the range of methods available for studying repair, extending the evidence base for its pervasive, systematic and universal character, and creating new points of contact between disciplines. The structure of repair lends itself to both manual and automated coding. However, judgements about the function of repair are more complex, context sensitive, and harder to incorporate into coding protocols. The process of defining repair protocols highlights the strengths and limitations of formal descriptions, the importance of choices about how data is captured and represented, and clarifies the relative importance of form, context and meaning for classifying different repair phenomena.

New methods for studying repair are producing evidence which can be difficult or impossible to detect in smaller samples, settings or populations. The capture and representation of multi-modal behaviour in repair sequences is challenging but novel forms of spatial data capture such as 3d motion capture are creating new opportunities and the accuracy and effectiveness of these methods will continue to improve. Despite their limitations, even somewhat noisy and approximate coding protocols can enrich what we know about repair and have wider payoffs, in particular enhancing the speed and scale of analysis and offering opportunities to investigate various phenomena such as the instrumental role of repair in natural versus task-oriented dialogue, detecting disorders such as Alzheimer's Disease and understanding the quality of clinical communication.

Acknowledgements

We are grateful to the editors and anonymous reviewers of this special issue for their insightful feedback that greatly improved this paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Patrick G.T. Healey http://orcid.org/0000-0003-3079-3374
Elif Ecem Özkan http://orcid.org/0000-0002-9036-2149

Bibliography

Andrews, D. (2014). Gestures as requests for information: Initiating repair operations in German native-speaker conversation. *focus on German Studies*, 21, 76–94.

Bain, M., Huh, J., Han, T., & Zisserman, A. (2023). Whisperx: Time-accurate speech transcription of long-form audio. *arXiv preprint arXiv:2303.00747*.

Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of personality and social psychology*, *79* (6), 941.

Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T., Zhang, F., & Grundmann, M. (2020). Blazepose: On-device real-time body pose tracking. arXiv preprint arXiv:2006.10204 .

Boersma, P. (2007). Praat: doing phonetics by computer. *http://www.praat.org/* .

Borges, N., Lindblom, L., Clarke, B., Gander, A., & Lowe, R. (2019). Classifying confusion: autodetection of communicative misunderstandings using facial action units. *Proceedings of the 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)* (pp. 401–406).

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech*, 44(2), 123–147.

Carletta, J. (1996). Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics*, *22*(2), 249–254.

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Clayman, S. E., & Heritage, J. (2025/this issue). Building and implementing coding systems in CA: An overview. *Research on Language and Social Interaction*.

Colman, M., & Healey, P. G.T. (2011). The distribution of repair in dialogue. In L. Carlson & T.F. Shipley (Eds.), *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*, Boston, Massachusetts (pp. 1563–1568).

Deppermann, A., & Gubina, A. (2025/this issue). Coding actions in social interaction: Potentials and problems. *Research on Language and Social Interaction*.

Dingemanse, M., & Enfield, N. J. (2015). Other-initiated repair across languages: towards a typology of conversational structures. *Open Linguistics*, *1*(1).

Dingemanse, M., Kendrick, K. H., & Enfield, N. J. (2016). A coding scheme for other-initiated repair across languages. *Open Linguistics*, *2*(1).

Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., … & Enfield, N. J. (2015). Universal principles in the repair of communication problems. *PloS one*, *10*(9).

Drew, P. (1997). Open class repair initiators in response to sequential sources of troubles in conversation. *Journal of Pragmatics*, 28(1), 69–101.

Floyd, S., Manrique, E., Rossi, G. & Torreira, F. (2016). Timing of Visual Bodily Behavior in Repair Sequences: Evidence From Three Languages, *Discourse Processes*, 53(3), 175–204.

Ginzburg, J. (2012). *The interactive stance*. Oxford University Press, USA.

Ginzburg, J., Fernández, R. M., & David, S. (2014). Disfluencies as intra-utterance dialogue moves. *Semantics and Pragmatics*, *7*(9), 64.

Gurion, T., Healey, P. G., & Hough, J. (2020). Comparing models of speakers' and listeners' head nods. *Proceedings of the 24th Workshop on the Semantics and Pragmatics of Dialogue. SEMDIAL*.

Healey, P. G. T. (2008). Interactive misalignment: The role of repair in the development of group sub-languages. In R. Cooper, & R. Kempson (Eds.), *Language in Flux: Relating Dialogue Coordination to Language Variation, Change and Evolution* (pp. 13–39). Palgrave-McMillan.

Healey, P. G. T., Colman, M., & Thirlwell, M. (2005). Analysing Multimodal Communication: Repair-Based Measures of Human Communicative Coordination. In J.C.J. van Kuppevelt, L. Dybkjaer, and N. Bernsen (Eds). *Advances in natural multimodal dialogue systems*, 113–129.

Healey, P. G. T., Lavelle, M., Howes, C., Battersby, S., & McCabe, R. (2013). How listeners respond to speaker's troubles. In M. Knauff, N. Sebanz, M. Pauen, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, Berlin, Germany (pp. 2506–2511).

Healey, P. G. T., Mills, G. J., Eshghi, A. & Howes, C. (2018). Running Repairs: Coordinating Meaning in Dialogue. *Top Cogn Sci*, 10, 367–388.

Healey, P. G. T., Purver, M., King, J., Ginzburg, J., & Mills, G. (2003). Experimenting with clarification in dialogue. In R. Alterman, & D. Kirsh (Eds.), *Proceedings of the 25th Annual Conference of the Cognitive Science Society*, Park Plaza Hotel, Boston (pp. 539–544). LEA.

Healey, P. G. T., Swoboda, N., Umata, I., & King, J. (2007). Graphical language games: Interactional constraints on representational form. *Cognitive Science*, 31, 285–309.

Healey, P. G. T., & Thirlwell, M. (2002). Analysing multi-modal communication: Repair-based measures of communicative co-ordination. *Proceedings of the International CLASS Workshop on Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems* (pp. 83–92).

Healey, P. T. G., Plant, N. J., Howes, C., & Lavelle, M. (2015). When words fail: Collaborative gestures during clarification dialogues. *Proceedings 2015 AAAI Spring Symposium Series*.

Hömke, P., Levinson, S. C., Emmendorfer, A. K., & Holler, J. (2025). Eyebrow movements as signals of communicative problems in human face-to-face interaction. *Royal Society Open Science*, *12*(3), 241632.
Hough, J. (2015). *Modelling Incremental Self-Repair Processing in Dialogue* (Doctoral dissertation, Queen Mary University of London).
Hough, J., De Ruiter, L., Betz, S. & Schlangen, D. (2015). Disfluency and laughter annotation in a light-weight dialogue mark-up protocol. *Proceedings of The 7th Workshop on Disfluency in Spontaneous Speech (DiSS)*, Edinburgh, UK, 49–52.
Hough, J., & Purver, M. (2014). Strongly incremental repair detection. *Proc. of Empirical Methods of Natural Language Processing (EMNLP)*. Doha, Qatar.
Hough, J., & Schlangen, D. (2017). Joint, incremental disfluency detection and utterance segmentation from speech. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, 1* (pp. 326–336).
Hough, J., Tian, Y., De Ruiter, L., Betz, S., Kousidis, S. Schlangen, D., & Ginzburg, J. (2016). Duel: A multi-lingual multimodal dialogue corpus for disfluency, exclamations and laughter. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 1784–1788
Howes, C., & Eshghi, A. (2021). Feedback relevance spaces: Interactional constraints on processing contexts in dynamic syntax. *Journal of Logic, Language and Information*, *30*(2), 331–362.
Howes, C., Hough, J., Purver, M., & McCabe, R. (2014). Helping, I mean assessing psychiatric communication: An application of incremental self-repair detection. Proceedings of the 18th Workshop on the Semantics and Pragmatics of Dialogue, Edinburgh, U.K.
Jefferson, G. (1972). Side sequences. p 294–451 In Jefferson, G., & Sudnow, D. (eds.). *Studies in social interaction*.
Jefferson, G. (1987). Exposed and embedded corrections. *Talk and social organisation*, 86–100.
Kendrick, K. H. (2015). Other-initiated repair in English. *Open Linguistics*, *1*(1).
Kitzinger, C. (2012). Repair. In Sidnell, J., & Stivers, T. (Eds.), *The handbook of conversation analysis* (pp. 229–256). John Wiley & Sons.
König, K., & Pfeiffer, M. (2025/this issue). Coding request for confirmation sequences: Methodological reflections from a cross-linguistic research project. *Research on Language and Social Interaction*.
Lake, J.K., Humphreys, K.R. & Cardy, S. (2011). Listener vs. speaker-oriented aspects of speech: Studying the disfluencies of individuals with autism spectrum disorders. *Psychon Bull Rev*, 18, 135–140.
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159–174.
Leudar, I., Thomas, P., and Johnston, M. (1992). Self repair in dialogues of schizophrenics: Effects of hallucinations and negative symptoms. *Brain and Language*, 43(3), 487–511.
Manrique, E. (2016). Other-initiated Repair in Argentine Sign Language. *Open Linguistics*. 2(1).
Manrique, E. & Enfield, N. (2015). Suspending the next turn as a form of repair initiation: evidence from Argentine Sign Language. *Frontiers in Psychology*. 6.
McCabe, R., Healey, P. G., Priebe, S., Lavelle, M., Dodwell, D., Laugharne, R., … & Bremner, S. (2013). Shared understanding in psychiatrist–patient communication: Association with treatment adherence in schizophrenia. Patient education and counseling, 93(1), 73–79.
McCabe, R., John, P., Dooley, J., Healey, P., Cushing, A., Kingdon, D., … & Priebe, S. (2016). Training to enhance psychiatrist communication with patients with psychosis (TEMPO): cluster randomised controlled trial. The British Journal of Psychiatry, 209(6), 517–524.
Mills, G., & Redeker, G. (2023). Self-Repair Increases Referential Coordination. *Cognitive Science*, *47*(8), e13329.
Mondada, L. (2016). Challenges of multimodality: Language and the body in social interaction. *Journal of sociolinguistics*, *20*(3), 336–366.
Mondada, L. (2018). Multiple temporalities of language and body in interaction: Challenges for transcribing multimodality. *Research on language and social interaction*, *51*(1), 85–106.
Mortensen, K. (2016), The body as a resource for other-initiation of repair: cupping the hand behind the ear. *Research on Language and Social Interaction*, 49(1), 34–57.
Nasreen, S., Rohanian, M., Hough, J., & Purver, M. (2021). Alzheimer's dementia recognition from spontaneous speech using disfluency and interactional features. Frontiers in Computer Science, 3, 640669.
Ozkan, E., Gurion, T., Hough, J., Healey, P. G. T., & Jamone, L. (2021). Specific hand motion patterns correlate to miscommunications during dyadic conversations, *IEEE International Conference on Development and Learning (ICDL2021), Beijing, China*.
Ozkan, E., Gurion, T., Hough, J., Healey, P. G. T., & Jamone, L. (2022). Speaker motion patterns during self-repairs in natural dialogue. *Companion Publication of the International Conference on Multimodal Interaction (ICMI '22 Companion)*. Association for Computing Machinery, New York, NY, USA, 24–29.
Ozkan, E., Healey, P. G. T., Gurion, T., Hough, J. & Jamone, L. (2023). Speakers raise their hands and head during self-repairs in dyadic conversations. *IEEE Transactions on Cognitive and Developmental Systems*, 15(4), 1993–2003.
Purver, M. (2002). Processing unknown words in a dialogue system. *Proceedings of the Third SIGdial Workshop on Discourse and Dialogue* (pp. 174–183).
Purver, M., Ginzburg, J., & Healey, P. (2003). On the means for clarification in dialogue. *Current and new directions in discourse and dialogue*, 235–255.

Purver, M., Hough, J., & Howes, C. (2018). Computational models of miscommunication phenomena. *Topics in cognitive science*, *10*(2), 425–451.
Purver, M. R. J. (2004). *The theory and use of clarification requests in dialogue* (Doctoral dissertation, University of London)
Randolph, J.J. (2005) Free-Marginal Multirater Kappa (Multirater K[Free]): An Alternative to Fleiss' Fixed-Marginal Multirater Kappa. *Proceedings of the Joensuu Learning and Instruction Symposium*, *Joensuu, Finland*.
Rasenberg, M., Pouw, W., Özyürek, A. & Dingemanse, M. (2022). The multimodal nature of communicative efficiency in social interaction. *Scientific reports*, 12(1), 19111.
Rasmussen, G. (2014). Inclined to better understanding—The coordination of talk and 'leaning forward' in doing repair. *Journal of Pragmatics*, *65*, 30–45.
Rodríguez, K. J., & Schlangen, D. (2004). Form, intonation and function of clarification requests in german task-oriented spoken dialogues. *Proceedings of Catalog (the 8th workshop on the semantics and pragmatics of dialogue (SemDial04)*.
Rohanian, M. (2023). Multimodal Assessment of Cognitive Decline: Applications in Alzheimer's Disease and Depression. (PhD Thesis, Queen Mary University of London)
Rohanian, M., & Hough, J. (2021). Best of both worlds: Making high accuracy non-incremental transformer-based disfluency detection incremental. *Proceedings of ACL-ICJNLP*, 1. 3693–3703
Rohanian, M., Hough, J., & Purver, M. (2021). Alzheimer's dementia recognition using acoustic, lexical, disfluency and speech pause features robust to noisy inputs. *Proc. of INTERSPEECH2021*, Brno, Czechia.
Sacks, H., & Jefferson, G. (1992). Harvey Sacks: lectures on conversation. *Malden*, *MA*: *Blackwell Publishing*.
Sacks, H., & Schegloff, E. A. (1979). Two preferences in the organization of reference to persons in conversation and their interaction. *Everyday language: Studies in ethnomethodology*, 15–21.
Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, *50*(4), 696–735.
Schegloff, E. A. (1987). Some Sources of Misunderstanding in Talk-in-Interaction. *Linguistics*, *25*(287), 201–218.
Schegloff, E. A. (1992). Repair after next turn: The last structurally provided defense of intersubjectivity in conversation. *American journal of sociology*, *97*(5), 1295–1345.
Schegloff, E. A. (1993). Reflections on quantification in the study of conversation, *Research on Language and Social Interaction*. 26(1). 99–128.
Schegloff, E. A. (1996). Some practices for referring to persons in talk-in-interaction: A partial sketch of a systematics. *Typological studies in language*, *33*, 437–486.
Schegloff, E. A. (1999). What next?: Language and social interaction study at the century's turn. Research on Language & Social Interaction, 32(1–2), 141–148.
Schegloff, E. A. (2000). On Granularity. *Annual Review of Sociology* 26, 715–20.
Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge University Press.
Schegloff, E. A. (2013). Ten operations in self-initiated, same-turn repair. In M. Hayashi, G. Raymond & J. Sidnell (Eds.), *Conversational repair and human understanding* (pp. 41–70). Cambridge University Press.
Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, *53*(2), 361–382.
Schiel, F. (1999). Automatic Phonetic Transcription of Non-Prompted Speech. *Proc. of the ICPhS*, 607–610, San Francisco.
Schlöder, J. J., & Fernández, R. (2014). Clarification requests on the level of uptake. *Proceedings of the 18th Workshop on the Semantics and Pragmatics of Dialogue (SemDial 2014, 'Dial-Watt')*, 237-239.
Schlöder, J. J., & Fernández, R. (2015). Clarifying intentions in dialogue: A corpus study. *Proceedings of the 11th International Conference on Computational Semantics*. 46–51.
Seo, M. S., & Koshik, I. (2010). A conversation analytic study of gestures that engender repair in ESL conversational tutoring. *Journal of Pragmatics*, 42(8), 2219–2239.
Shriberg, E. (1994). *Preliminaries to a theory of speech disfluencies* (Doctoral dissertation, University of California, Berkeley).
Stivers, T., & Enfield, N. J. (2010). A coding scheme for question-response sequences in conversation. *Journal of pragmatics*, *42*, 2620–2626.
Stivers. T., & Rossi, G. (2025/this issue). Finding Codability: Ways to code and quantify interaction for Conversation Analysts. *Research on Language and Social Interaction*.
Themistocleous, M., McCabe, R., Rees, N., Hassan, I., Healey, P. G. T., & Priebe, S. (2009). Establishing mutual understanding in interaction: An analysis of conversational repair in psychiatric consultations. Communication & medicine, 6(2), 165–176.
Wilkinson, S., & Kitzinger, C. (2006). Surprise As an Interactional Achievement: Reaction Tokens in Conversation. *Social Psychology Quarterly*, 69(2), 150–182.
Wilkinson, S., & Weatherall, A. (2011). *Insertion repair*. Research on Language and Social Interaction, 44(1), 65–91.
Wittenburg, P., Brugman, H., Russel, A., Klassmann, A. & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. *In 5th international conference on language resources and evaluation (LREC2006)* (pp. 1556–1559).