

How are translations created? Using multimodal conversation analysis to study a team translation process

Maija Hirvonen

University of Helsinki, Finland
mai.hirvonen@helsinki.fi

Liisa Tiittula

University of Helsinki, Finland
liisa.tiittula@helsinki.fi

This article demonstrates a methodology for studying the translation process from the perspective of multimodal social interaction and applies this methodology to a case analysis of collaborative audio description. The methodology is multimodal conversation analysis, which aims to uncover the way in which multimodal communication resources (e.g., talk, gaze, gestures) are used holistically and situatedly in building human action. Being empirical and data-driven, multimodal conversation analysis observes human conduct in its natural setting. This article analyses video data from an authentic audio-description process and presents the multimodal constitution of problem-solving sequences during translating. In addition, the article discusses issues regarding the methodological choices facing researchers who are interested in human interaction in translation. The article shows that applying multimodal conversation analysis opens new avenues for research into the translation process and collaborative translation.

1. Introduction: From textual to social-interactive

Taking a multimodal perspective on translation often means that multimodality is considered a *textual* phenomenon, that is, as the meaning-making potential of visual, auditory and verbal elements in both source texts (ST) and target texts (TT) (see, for example, Kaindl, 2013 and articles in this Special Issue). In research into interpreting, on the other hand, multimodality becomes a *social-interactive* phenomenon in interpreter-mediated interaction where, in addition to talk, bodily actions and material artefacts used by co-participants are analysed (see, for example, Davitti & Pasquandrea, 2017). In the present article, we show how multimodality constitutes collaborative translation processes.

The goal of the present article is to demonstrate a methodology that takes a conversation-analytical approach to multimodal interaction and is also suitable for studying multimodal interaction in translation processes. We sample the method using an analysis of team translation in audio description. Multimodal conversation analysis (CA) is widely used in social and linguistic interaction research. It examines the orderliness of human interactions from a multimodal perspective, therefore analysing people's use of both verbal (talk) and embodied communication (gestures, facial expressions, posture, etc.) and material objects on the basis of video data of naturally occurring social interactions. Accordingly, multimodality includes all the relevant resources employed by the participants to build and interpret their situated action and it treats them in the same way without *a priori* hierarchization (Mondada, 2018, p. 86).

The case we analyse involves a team of three translators scripting an audio description for a film; one of them is blind and two are sighted. Although not commonplace globally, team translation is used in the audio description praxis for film and television in Germany and, with some variations, in Austria and Finland. Despite this relative marginality, studying team audio description in the framework of multimodal CA enables one to “zoom in” to the micro level of the translation process in order to answer the questions: (1) How are problems solved and decisions made in teams?; and (2) How are different resources being used in translating?

We begin with a brief rationale for our approach and case study in relation to the research on the translation process, interaction and collaboration. Then we present the principles of multimodal CA. This includes the introduction of essential theoretical standpoints as well as the methodological implications for data-gathering and data management. After that, we present a sample analysis of team audio description. The reader will notice that the analysis of multimodal interaction takes a great deal of space, even for brief data samples. Finally, we summarize the findings of our sample analysis and discuss the implications of choosing this methodology over multimodal translation process research.

2. Rationale

The question we pose in the title—how are translations created?—involves a process-oriented perspective on the study of translation (see Kussmaul, 1995). Processes can be studied with different methodological approaches. Traditional methodologies focus on the performance of individual translators, while more recent approaches include collaboration as a relevant feature of the translation process. While “the cognitive processes involved in performing a translation task are not available for direct observation” (Englund Dimitrova, 2010, p. 407), collaboration as social interaction between translators is.

The individual cognitive processes of translation have largely been studied in experimental settings. The introspective method can be used either in the course of translating or retrospectively after the process. The most frequently used introspective method is thinking aloud, in which the subjects verbalize their thoughts while translating the source text, the verbalizations are recorded, and then they are transcribed in think-aloud protocols (e.g., Jääskeläinen, 1999). The method has been criticized as unnatural; Kussmaul (1995), for example, has suggested dialogue protocols in addition to monologue protocols involving a pair of subjects instead of a single individual to carry out a translation task. Pavlović (2013) proposes the term “collaborative translation protocols” (CTPs) and claims that these are not think-aloud protocols in the strict sense because they also include interaction. Pavlović’s study aimed at comparing collaborative and individual translation and focused on problems encountered in both tasks, not on the interactive action of collaboration. Further, frequently used methods in process research are key-stroke logging and eye-tracking, which enable a close examination of the writing process (e.g., Göpferich et al., 2008; Hansen, 2006; see also Englund Dimitrova, 2010; O’Brien, 2011a). One challenge in all of these methods is their ecological validity: the participants are not carrying out their usual translation tasks in their normal environment. Also, the participants typically consist of students.

On the social and interactive side, the translation process is investigated as happening in the real working context in the framework of embodied/embedded or extended cognition, which sees cognition as “a context-dependent interaction of mind, body and environment” (Risku & Windhager, 2015, p. 37; see also Muñoz-Martín, 2017). The study of situated translation in practice includes the observation of everyday working life and the analysis of interactions between people and their social and material environments (e.g., Risku & Windhager, 2015). The translation process can also be

understood by analysing versions of translations and related archive material. Walker (2017), for instance, examines the contributions and interactions of individuals involved in the production process of a translation (e.g., editor, author and translator). His study is based on an analysis of postal correspondence between these individuals.

All in all, collaboration between different translation agents has been attracting interest in translation studies (see, for example, Cordingley & Frigau Manning, 2017). A variety of research areas are currently examining translation as a joint achievement of translators, revisers, editors, clients and commissioners or of human–machine interactions (O’Brien, 2011b). This research ranges from literary translation (Walker, 2017) to the present-day interpreting (Chmiel, 2008) and online translating (Jiménez-Crespo, 2017). We complement the existing research into process and collaboration in translation in two ways. Instead of focusing on translation *events*, our interest lies in the *act* of translating (cf. Chesterman, 2015). While the act has so far been the object of experimental research, with a concentration on the mental processing and the performance of individual translators, in our research the analysis of the ST and the formulation of the TT become interactively displayed and intersubjectively shared.

3. Conversation-analytical approach to multimodal interaction

3.1 Principles of CA-based multimodal interaction analysis

CA is not only a method with which to analyse talk, but it is also a theory of social interaction. It is rooted in sociology and was developed in the 1960s and late 1970s by Harvey Sacks and his collaborators, Emanuel Schegloff and Gail Jefferson. CA emerged out of ethno-methodology and its analysis of “folk” (ethno) methods, that is, how collective members create and maintain a sense of order and intelligibility in social life (ten Have, 2004, p. 14).

CA is concerned with social action in interactions. The basic assumption is that social action is structurally organized and deeply ordered. This orderliness is studied as a mutual achievement of the actors. The research has focused on talk, since talk has been regarded as the primordial site of interaction and the main medium through which the interactants construct a sense of the ongoing event. Through talk, they create and sustain intersubjective understanding. CA is not concerned with the internal processes of what goes on in the mind, but rather with what is directly observable: the practices used by the interactants to construct understanding and the development of intersubjectivity in an action sequence. (For overviews of CA, see, for example, ten Have, 2007; Sidnell & Stivers, 2013.)

The basic organization mechanics constitutive of talk-in-interaction are turn-taking, sequence organization, and repair. Turn-taking organization concerns the relative ordering of speakers and the construction of turns. Sequence organization refers to any kind of organization that involves the relative positioning of utterances or actions. Repair organization refers to the practices of dealing with problems in the interaction (Schegloff, 2007, p. 2).

Actions are simultaneously context shaped and context renewing. This means that they are understandable as reactions to the co-participant’s prior action and, at the same time, they project a certain type of subsequent action. The next turn displays the interactant’s understanding and interpretation of the prior turn. A central format of a sequence is the adjacency pair (e.g., question–answer, greeting–greeting). The pair is composed of two different turns by different speakers, is relatively ordered and is pair-type related (Schegloff, 2007, p. 13). In pairs such as invitation–acceptance/refusal, the alternative responses are not symmetrical, but construct a preference organization. The

dispreferred second pair part, such as a refusal, requires more conversational effort than the preferred one and is typically produced with delay, mitigation, and elaboration. The preference order is an interactional feature, not a psychological one; the participants display their orientation to it in the way they produce their response: either straightforwardly or elaborated (Schegloff, 2007).

CA is strictly empirical and data-driven. Situated activities are observed in their natural setting; basic data are audio or video recordings, which allow the interaction to be observed repeatedly. The recordings are transcribed in order to be able to analyse the data in detail.

At present, CA researchers are increasingly interested in the analysis of multimodality in interaction, although video-based multimodal research has a long history in workplace studies, in which collaborative activities in everyday workplace settings are analysed (e.g., Goodwin & Goodwin, 1996; Luff et al., 2000). Analysis of the different audible and visible resources that people use as well as analysis of the material objects and surroundings are required, since CA aims at a comprehensive understanding of human interaction (Deppermann, 2013, p. 2) and of how multimodal resources are used holistically and situatedly in building human action (Mondada, 2018).

In addition to Sociology, the methodology of CA has been used in Linguistics, Communication Studies and many other disciplines. In Translation Studies, the research on dialogue interpreting has increased interest in interaction and CA-based multimodal analysis (e.g., Davitti & Pasquandrea, 2017).

3.2 Data-gathering and -analysis in practice

CA examines only naturally occurring data and does not accept material from experimental settings. The reason for this is the effect of unnatural behaviour that experiments can have on participants' conduct. Data are therefore gathered on site, where people meet and interact in everyday settings. Since everything in the interaction, including non-verbal behaviour, is potentially relevant to the analysis, the interaction has to be recorded. In the early phases of CA, much work focused on telephone conversations using audio recordings, but now video recordings have become common and visual aspects have gained more attention.

Heath et al. (2010) provide a practical guide to video-based research and address many problems involved in collecting, analysing, and presenting video data such as ethical issues, the observer's role, including camera position and the use of multiple cameras, and the equipment used.

Since human beings are the research subjects of CA, gaining access to make video recordings is essential. This includes obtaining permissions to record and use the data. Informed consent is usually given in written form. In using the data, researchers must comply with measures to preserve the privacy of the participants and other parties involved. This means, for instance, that data are anonymized.

The camera and the observer-researcher always have a certain viewpoint, and the choices made in the set-up of data-gathering have implications for the analysis. When a fixed camera is used, the viewpoint has to be selected prior to beginning a recording session. This has led to a preference for using multiple cameras, especially when several participants are involved and the activities in the setting are dispersed. The quality of the video and audio needs to be good. For instance, the use of an external microphone is recommended, otherwise it can be difficult or even impossible to analyse the talk of the participants.

The data, talk and visible conduct have to be transcribed for systematic analysis. A detailed transcript captures not only verbal utterances, but also prosodic features, voice quality, emphasis, timing and laughter as well as such interactive phenomena as overlapping. The transcription system developed by Gail Jefferson is prevalent in most

CA studies. For multimodal analysis, particular conventions have been developed to annotate embodied actions, such as gesture, gaze, posture, body movements, and so on that happen simultaneously during talk or in moments when there is an absence of talk (see, e.g., Luff & Heath, 2015; Mondada, 2018). Often, still images are used to complement the “traditional” transcription.

Transcripts are necessary for close scrutiny, but since all transcriptions are selective, the analysis has to be based on the recordings. CA is used to examine the practices used by the interactants to accomplish actions. Special attention is paid to how they manage this in a sequential way, step by step, through turns-at-talk. In the framework of CA, only features that are made relevant by the participants themselves in the interaction are taken into account. Therefore, the background information of the participants (including their age, educational background, work experience, and so on) is typically not reported and it does not inform the analysis of the interaction. The analysis begins with single cases, but after identifying an instance of special interest, the analyst tries to gather a collection of the same kinds of instance in order to find regularities.

4. Applying the methodology: Face-to-face team translation

Our case study¹ deals with a real-life “collaborative translation task” (cf. Pavlović, 2013), which is a type of writing-in-interaction (Mondada & Svinhufvud, 2016): a team of translators meets face-to-face to translate a text together, from start to finish. In our case, three audio describers (Lisa, Ines and Sara, who is blind) sit and work together to write a script for a film audio description, a task that consists of several phases and subtasks. The main phases can be identified as (1) the planning, orientation, and reading phase (cf. Englund Dimitrova, 2010, p. 409), where the team watches and discusses the filmic source material together, and the sighted team members, Lisa and Ines, describe the visuals to Sara *prima vista*; and (2) the drafting/TT-generating phase, in which a first draft of the script is prepared. The subtasks of the second phase include the selection of relevant information to be described (what to describe), the formulation of the descriptions (how to describe) and spotting (where to describe). As regards the use of tools, Ines manages the laptop on which the source film² is played and Lisa types the script on another laptop. The team does not have other material about the film that could support the audio description, such as a dialogue list or a film script.

Our interest lies in uncovering translating as an interactive practice: how a particular target-text segment is created in the interaction between the audio describers. We analyse interaction from the multimodal perspective—thus not merely as a verbal phenomenon expressed in the spoken utterances of the participants, but also how the translation is produced through embodied resources (gaze, gestures, posture, etc.) and how the comprehension of the meanings of words is shared and negotiated and mutual understanding is established.

In what follows, we demonstrate a sample analysis of multimodal interaction in team translation according to a typical CA procedure: we begin with a description of the dataset, continue with a step-by-step analysis of the sequential constitution of interaction, and end with a summary of the findings.

4.1 Data

To demonstrate the method, we have chosen two sequences from a dataset which in total consists of two working days of video-recorded team translation. These two sequences (the first is 13 seconds long and the second is 70 seconds) illustrate a central phase of audio description and, in fact, of translating as a whole: the interactive and multimodal

solving of a problem and making of a decision. This concerns mainly two subtasks, which are both negotiated by the team members: (1) WHAT visual information will be described; and (2) HOW that visual information will be described. Our analysis aims to find out how the second subtask is performed during the social interaction of a translating team. We are interested in the meaning negotiation that precedes the actual typing of the script, and to this end we analyse two sequences of this action.

The transcript was created on the basis of a one-camera video and the film copy. The transcription conventions are presented at the end of this article. For the analysis, we have shown in bold the target-text words that are being negotiated. We use abbreviations to refer to the different speakers and sources of information: FS = film sound, FD = film dialogue, IN = Ines, LI = Lisa, SA = Sara, ? = unidentified speaker.

4.2 Analysis: Solving a problem in translating

We join the team at a moment when it starts to describe a new segment of a film scene. Ines (on the left of the image) and Lisa (in the centre) are oriented visually towards the film. Sara (on the right) sits back with her arms crossed across her lap. She stares downward, seemingly with her eyes closed, but the blinking of her eyelids shows that she is listening and is therefore auditorily oriented towards the film. Ines describes the film scene as the film plays.

Sequence 1:



Figure 1 (corresponds to the transcription line 2 below)

```

1 FS: [(street noise, 8.0) ]
2 IN: [(mumbles °(-)° ) (.) drinnen (.) in einem flur]
           inside in a hallway
3 FD: (male voice) ihre mutter [ist sei- ]
           your mother is sinc-
4 IN: [hapapaa ]

```



Figure 2 (corresponds to the transcription line 5 below)

5 LI: =in einem **gang** ne?
*in a **passageway** right*

6 IN: =in einem **gang**;
*in a **passageway***

7 ?: °mhm, °

8 (2.0)

9 LI: (-) (sogar) (.) >machen wir erst mal< ö:
even let's do first uhm

10 nee ich würd' gar nicht drin machen
nope I wouldn't actually make it inside

11 sondern in einem **gang** (...)
*but in a **passageway***

Ines describes the film scene while simultaneously playing the film: while street noise is audible (line 1), Ines mumbles something very quietly, which is unfortunately unintelligible³ (line 2). Then Ines utters more clearly *drinnen* (.) *in einem Flur* “inside in a hallway”. The utterance is marked as uncertain (with pauses and a slightly rising intonation) and negotiable, thus inviting a confirmation or disconfirmation from the co-participants. Lisa is attentive to Ines and the film, gazing at the film screen and leaning toward Ines (Figure 1). Ines makes her utterance relevant to Lisa multimodally (see Mondada & Svinhufvud, 2016, p. 27) with a slightly rising intonation, by pausing the film, by shifting her gaze toward Lisa’s laptop, and finally by turning her head toward Lisa (Figure 2). This prompts a response from Lisa, who rejects the lexical choice *Flur* “hallway” and replaces it with *Gang* “passageway”. Lisa finalizes her utterance with a negative tag question *ne* “no” in a rising intonation (line 5). This kind of turn design seeks agreement from the recipients. Ines immediately reacts, repeating Lisa’s formulation with slightly falling intonation, which marks acceptance of this choice (line 6). She also shifts her posture back to the earlier position, displaying that the problem is solved. The proposal is also approved by an interjection *mhm* (line 7), although we cannot recognize from the video who says it. The mutual agreement of “passageway” as the correct term at this point is further displayed by the fact that Lisa begins to use it, with an emphatic prosodic stress, in her description proposal (line 11). A few seconds later, she also types it (*in einem Gang* “in a passageway”) in the script, exhibiting with the writing that the proposal has been accepted (cf. Mondada & Svinhufvud, 2016, p. 25).

Sequence 2:

Thirteen minutes later, the team has progressed to a later part in the same scene. As earlier, the team has just watched and listened to the film; Ines has stopped the film, and now Lisa verbalizes the visuals to Sara:



Figure 3 (corresponds to the transcription line 14 below)

12 LI: jetzt: also >das=is köh ↑licht also das fahle licht
now so that's uhm light the pale light

13 ähm< (0.4)
uhm

14 ALSO ES GIBT SO EIN'N GRELLEN STREIFEN
so there is like a bright stripe

15 AM ENDE DES GANGES DAS=IS NÄMLICH EIN FENSTER;
at the end of the passageway it is actually a window

16 UND DA FÄLLT SO N GRELLER STREIFEN ÄH
and there comes like a bright stripe uh

17 TAGESLICHT DREIN DER SICH DANN ABER SO: (.)
daylight in which then like

18 VERWÄSCHT (0.3) [JE W]EITER DER GANG IS.
blurs away the longer/wider the passageway gets

19 SA: [mhm,]

Lisa's prosodic and bodily action is directed at Sara. She pronounces her turn (lines 14–18) louder than the surrounding talk, and her body is inclined towards Sara (Figure 3). Her gaze is mostly directed at the film, but she looks at Sara twice during her turn and glances quickly at Ines during the second mention of “bright stripe” (16). Although focused here on describing the atmosphere—the lighting—Lisa uses the word *Gang* “passageway” twice (15 and 18). On both occasions, she attributes certain qualities to it: first, she verbalizes that its end is visible by using the prepositional phrase *am Ende des Ganges* “at the end of the passageway” (15), which serves to locate a particular ST item, “bright stripe”, in a certain place in the setting. Second, Lisa attributes a physical property to the passageway with the comparative form *je weiter der Gang ist* “the longer/wider the passageway gets” (18). This description prompts a minimal response, a continuer, from Sara (line 19) with a continuing intonation, which indicates that she is following the talk, but has nothing to add or contradict at this point. A short pause follows, after which Lisa continues:



Figure 4 (corresponds to the transcription line 25 below)

20 (0.6)
 21 LI: vielleicht kann man kurz die stimmung reinnehmen
perhaps we can briefly include the atmosphere
 22 von dem flur dann geht sie in dies- zimmer, *sniffs*
from the hallway then she goes in the room
 23 IN: (°ja dass es so eng ist°)
yes that it is so narrow
 24 LI: ja (.) >ich hab auch gedacht<
yeah I also thought
 25 ob es dann doch nicht eher ein flur wäre?
wouldn't it actually rather be a hallway
 26 öhm wie du vorher schon sagtest; (0.5)
uhm like you already said before
 27 weil der so gesch==schmal is;
because it is so (x)- narrow

Now Lisa addresses Ines (line 21): she turns her head towards the video and Ines (the turning of head is also audible in the change of voice quality) and lowers her voice (Ines sits closer to her than to Sara). She now describes the location with the word *Flur* “hallway” instead of *Gang* “passageway” and by so doing returns to the first candidate, introduced in Sequence 1. Her proposal prompts a response by Ines (23), who first acknowledges the proposal with the response particle *ja* “yes” and then offers her understanding of what is characteristic of the location: *dass es so eng ist* “that it is so narrow”. Again, Ines’s speech is very quiet and therefore difficult to discern, but at least the adjective *eng* “narrow” comes out. She accompanies her utterance with an iconic gesture visualizing NARROWNESS.

Lisa acknowledges Ines’s observation about narrowness with “yes” (24) and elaborates her turn with a string of utterances. First, she verbalizes her thinking process: “I also thought wouldn’t it actually rather be a hallway” (24–25). Then she expands her turn with a string of subordinate clauses, which justify her proposal further and refer to

the earlier discussion “like you said before” (26) and to a physical property “because it is so (x)- narrow” (27). While uttering this, Lisa is in a “thinking posture” (Figure 4), with her hand covering her mouth and arms tightly across her torso. As she utters *weil ...*, she turns her head and gaze to Sara, thus selecting her as the next speaker. This movement produces a change in the direction of her voice, thus making it acoustically perceptible to Sara as well to Ines. Sara then takes the next turn.



Figure 5 (corresponds to the transcription line 32)

- 28 SA: mhm: mhm, [*click of tongue*]
 29 LI: [also bei **gang** assoziiere ich immer
so with passageway I always associate
 30 etwas **breite**[res:] so wie im krankenhaus,
something wider like in a hospital
 31 [und bei **flur**, ähm assoziiere=ich] was **schmales**.
*and with **hallway** uhm I associate something narrow*
 32 IN: [(das ist) fast so wie bei einem schiff hier]]
it's almost like in a ship here

Sara reacts with two response particles (line 28), the first of which entails a stretch of the final sound (*mhm*) and the second ends with continuing intonation. Sara is then prepared to take the turn with a click of the tongue and by opening her mouth, but Lisa continues with another justification of her willingness to change the word (29–31). In a body-torque posture (Schegloff, 1998), Lisa projects two actions at the same time: her head and gaze address Ines and her torso remains facing her own laptop. She shares her associations of the candidate words: “with *passageway* I always associate something wider like in a hospital and with *hallway* uhm I associate something narrow” (29–31). Overlapping with Lisa’s last utterance, Ines makes a comparison “it’s almost like in a ship here” (32) and repeats her gesturing action (hands imitating narrowness, Figure 5). She does not, however, offer any particular term. Ines accompanies her utterance with a smile and continues to smile during the others’ turns.



Figure 6 (corresponds to the transcription line 33 below)

- 33 LI: (.) .h wie geht es dir?
what about you
- 34 SA: ja hast du irgendwie recht wobei
yeah you are somehow right although
- 35 ich bei **flur** trotzdem immer ein bisschen
anyway with hallway I always associate a little bit
- 36 also was gemütl- in anführungs(zeichen)
I mean something cos- in quotation(marks)
- 37 gemütlicheres weil
cozier because
- 38 =[weil mit wohnung verbunden, (ich) denke.]
since [it's] connected with apartment I think
- 39 [(swishing sound)]
- 40 LI: mhm̩
- 41 (1.7)

After explaining her association of hallway with the adjective *schmal* “narrow”, Lisa turns again to Sara, addressing her with a gaze as well as by the direct question of her opinion (line 28 and Figure 6). In the body-torque position, Lisa’s torso is oriented towards the film for analysis and her head is turned to Sara for negotiation. Ines, too, looks at Sara and smiles. Sara takes her turn and begins to elaborate her thoughts (34–38). Sara begins her response with a “pro forma agreement”, which delays her differing opinion (Schegloff, 2007, pp. 69–70). Her utterance displays hesitancy (self-repair, repetition) and is mitigated (“somehow”, “a little bit”, “in quotation marks”, “I think”). Disagreements are dispreferred second pair parts, which are typically produced with these kinds of modification. During Sara’s turn, Lisa gradually alters her attentive posture of gazing at Sara by shifting her gaze away (starting at *bei Flur*, line 35) and by leaning

sideways to grab a water bottle at her feet. Ines, although silent, is monitoring the interaction and looking at Sara as she speaks, her eyes narrowing when Sara uses the expression “in quotation (marks)” (36). Ines shifts her gaze to the film, while Lisa, in contrast, looks at Sara as she justifies her disagreement (38). Sara gazes towards Lisa. Lisa opens the water bottle, which produces a swishing sound simultaneously with Sara’s offset (39). Lisa responds with a nod and a minimal token *mhm* with slightly rising intonation, which indicates information received but does not close the problem-solving sequence. Lisa turns her gaze back to the film. Then, for a relatively long (1.7 seconds), the discussion pauses. Lisa and Ines look at the film, Ines is in a thinking posture, and Lisa closes the bottle so as to project the next action. Sara sits with her arms folded and leans back in her seat, gazing straight ahead. The pause ends with Lisa’s verbal action toward Ines:



Figure 7 (corresponds to the transcription line 44 below)

42 LI: **korridor** hatten wir auch schon mal
*we already had **corridor***
 43 bei (de)m durchgang gehabt
in the walkthrough
 44 oder war das was anderes [oder]
or was that something else or
 45 SA: [**korridor**]
corridor
 46 ist ganz breit eigentlich.
is actually quite wide
 47 (0.4)

48 LI: jaꝫ
really

49 (1.3)

50 ?: (müssen [wir (-)]
we must

51 SA: [war nur meine(r)] assoziation jetzt hehe
was only my association now ha ha

52 ((two clicks, 0.8))

53 SA: [also du kannst,]
so you can

54 LI: [okay du sagst] **korridor** nicht gut
okay you say corridor not good

55 weil du sagst breit, [na gut]
because you say broad alright

56 [((CLICK))]

57 LI: dann lassen wir gang.
then we leave passage

58 (0.8)

59 SA: hmh;

60 (0.6)

61 LI: ähm: genau. .h >ABER WEIL'S JETZT RELATIV
uhm right because it now takes relatively

62 LANG[E DA]UERT BIS SIE REINGEHTꝫ<
long before she goes in

63 SA: [mhm,]

64 LI: hab ich gedacht wär's ganz schön:
I thought it'd be quite nice

65 (2.3)

Now Lisa makes a new proposal and offers the word *Korridor* “corridor”. However, she downgrades it immediately as having already been used for another purpose. Lisa asks the co-participants for confirmation, yet bodily and visually she is directing her remark to Ines (line 44, Figure 7). Ines remains immobile but, without delay and overlapping the turn-yielding particle *oder* “or”, Sara takes the turn and confirms Lisa’s assessment, attributing to the word the feature WIDE (46). Sara also turns her gaze toward Lisa. The particle *eigentlich* at the end of her utterance mitigates her statement and her possible disagreement with Lisa’s proposal. Lisa’s delayed minimal response with rising intonation (48) signals her disagreement with the proposal. This is exactly Sara’s interpretation, as her next turn shows: she further mitigates her statement and produces laugh particles (51). Laughter typically occurs in situations of embarrassment and discomfort, including in dispreferred structures. With laughter, the speaker can mark her talk as laughable (Glenn, 2003, p. 48) and thus modify her action. The pauses as well as Lisa’s action of drinking water, which is secondary to the translation task, also signal trouble in the interaction and the difficulty of the task at hand. Sara begins to utter a proposal (53), and her arms begin to unravel from their crossed position. Yet overlapping her onset, Lisa acknowledges Sara’s assessment (“okay”) and quotes Sara’s assessment, reformulating it as evaluation (“not good because”) (54–55). Then she claims the word

choice *Gang* “passageway” as the solution, with briefly and slightly rising shoulders. Her turn is followed by a gap until Sara produces an acknowledgment token (59). After a second gap, Lisa again confirms the choice, but begins to explain her description proposal, which in fact takes the discussion back to the point before the word choice was claimed as problematic. The initial particle *aber* “but” (61) projects a contrast to what was said.

4.3 Findings of the sample analysis

We now summarize the most relevant findings of our analysis. Translating—here, audio describing—proceeds from identifying a problem (with a possible explanation) to proposing and negotiating candidate solutions and, finally, to solving the problem by reaching a decision that is mutually acceptable. The problem is solved when someone proposes a solution and the others either affirm it or do not initiate new proposal sequences. Such problem-solving sequences present cases of proposal/acceptance and proposal/negotiation/acceptance sequences, which are found in other situations of writing-in-interaction (Mondada & Svinhufvud, 2016, p. 25).

In this procedure, actions are performed multimodally. A translation problem, for instance, can be made relevant to others by marking a discussion item grammatically (e.g., with an interrogative) or prosodically (with rising intonation), while simultaneously using tools (e.g., gazing or pointing at laptops on which the source film and the target script are used) and one’s own body (e.g., depicting one’s conceptualization with an iconic gesture). The multimodal analysis also shows that involvement is not necessarily verbal, but co-participants actively monitor and therefore participate in the meaning negotiation with their embodied resources, such as gaze. Word meanings may be presented to others partly non-verbally, such as by gesturing and describing some properties of the item in the ST, but without ever actually proposing an actual translation for it. Accordingly, the deverbalization stage (see Lederer, 2010) becomes observable as translating is embodied.

In the analysed sample, co-participants justify their candidate suggestions based on the associations that these words evoke in them. Our analysis thus consolidates the assumption by Seiffert (2005) and Hirvonen and Tiittula (2012), made on the basis of textual analyses of audio description, that schematic knowledge is a decisive factor in problem-solving in audio description. In the case of choosing an adequate concept to describe the location in the film, that is, whether to call the location a hallway, a corridor or a passageway, the decisive associations relate to physical properties and to the schematic knowledge related to the concept (e.g., a “hallway” of an apartment, and therefore a pleasant atmosphere or not). Multimodal interaction analysis therefore enables the way in which translating unfolds to be investigated as a process prior to the final text.

5. Discussion

In this article, we have applied multimodal CA to study the translation process from the perspective of social interaction among individuals. Even though research on translation processes and collaborative translation is flourishing, it has so far overlooked the detailed analysis of face-to-face interaction among translators in their natural settings. We have shown how multimodal CA makes such interaction directly observable. Although here we have sampled an analysis of a specific type of translating, we claim that this methodology is also applicable to more typical cases of interlingual translation—for example, to study the settings in which translators (and revisers) meet to discuss drafts of translations.

Similar to the multimodal turn in text analysis, the multimodal analysis of interaction has widened the focus from talk to the way people use their bodies and environments as communication resources. Our analysis shows how collaborative translating can be based on multimodality. The relevance of the multimodal perspective is further demonstrated by the observation that involvement in the translation process can be “silent”, which would not be analysable on the basis of talk alone. Even if a translator is not speaking or actively writing, this does not “pause” action; rather, they can be involved in monitoring the ongoing activity or projecting the next one. This relates to a common challenge in translation process research, namely, to investigate what happens in translation during pauses or gaps when translating does not leave a trace. Methodological triangulation, such as the combination of key-logging with eye-tracking, is often mentioned as providing access to this complex processing (Göpferich, 2008, p. 51).

Our analysis produced a data-driven account of real-life problem-solving in translation. The findings of such analyses, when these are carried out on larger collections of cases, can be compared to theoretical models and other types of research on translation processes (e.g., Shih, 2015). While experimental research intends to minimize the effect of the “noise” of natural settings, the goal of multimodal CA is precisely the opposite: it reveals the nuanced constitution of the social interaction of translators and shows how interactive phenomena, such as “rationalization” as meaning negotiation and “interpersonal relations” as mutual understanding (cf. Pavlović, 2013, p. 552), are constituent elements of translation.

The case we have studied is also an example of meaning negotiation, that is, the “process through which agents ... converge to an agreement through some communication medium” (Warglien & Gärdenfors, 2015, p. 80). Intersubjective meaning negotiation offers a way of dealing with the objectivity versus subjectivity debate and opens new avenues for research into translation.

Transcription conventions

(0.4)	A pause and its duration (seconds)
(.)	A micropause (less than 0.2 seconds)
=	Latching, i.e., no interval between talk segments
[Beginning of overlap
]	End of overlap
.	Falling intonation
;	Slightly falling intonation
,	Continuing intonation
?	Rising intonation
;	Slightly rising intonation
<u>ein</u>	Uttered with emphatic stress
↑	Uttered with a higher pitch than the surrounding talk
°eng°	Silently pronounced word or utterance
ALSO	Pronounced more loudly than the surrounding talk
>ja<	Uttered faster than the surrounding talk
gesch-	A cut-off word (self-interruption)
mhm:	Lengthening of a sound
schö:n	Binding of sounds together
.h	Inbreath
(eng)	Transcriber’s doubt
(-)	An unclearly heard word or utterance
(...)	Omitted stretch of talk
sniffs	Non-verbal action

References

- Chesterman, A. (2015). Models of what processes? In M. Ehrensberger-Dow, B. Englund Dimitrova, S. Hubacher-Davidson, & U. Norberg (Eds.), *Describing cognitive processes in translation: Acts and events* (pp. 7–20). Amsterdam: John Benjamins. doi:10.1075/bct.77.02che
- Chmiel, A. (2008). Boothmates forever?: On teamwork in a simultaneous interpreting booth. *Across Languages and Cultures*, 9(2), 261–276. doi:10.1556/Acr.9.2008.2.6
- Cordingley, A., & Frigau Manning, C. (Eds.). (2017). *Collaborative translation: From the Renaissance to the digital age*. London: Bloomsbury.
- Davitti, E., & Pasquandrea, S. (2017). Embodied participation: What multimodal analysis can tell us about interpreter-mediated encounters in pedagogical settings. *Journal of Pragmatics*, 107, 105–128. doi:10.1016/j.pragma.2016.04.008
- Deppermann, A. (2013). Multimodal interaction from a conversation analytic perspective. *Journal of Pragmatics*, 46, 1–7. doi:10.1016/j.pragma.2012.11.014
- Englund Dimitrova, B. (2010). Translation process research. In *Handbook of translation studies*, Volume 1 (pp. 406–411). Amsterdam: John Benjamins. doi.org/10.1075/hts.1.tra6
- Glenn, P. (2003). *Laughter in interaction*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511519888
- Goodwin, C., & Goodwin, M. H. (1996). Seeing as a situated activity: Formulating planes. In Y. Engeström & D. Middleton (Eds.), *Cognition and communication at work* (pp. 61–95). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139174077.004
- Göpferich, S. (2008). *Translationsprozessforschung: Stand, Methoden, Perspektiven*. Tübingen: Narr.
- Göpferich, S., Jacobsen, A. L., & Mees, I. M. (2008). *Looking at eyes: Eye-tracking studies of reading and translation processing*. Frederiksberg: Samfundslitteratur.
- Hansen, G. (2006). *Erfolgreich Übersetzen: Entdecken und Beheben von Störquellen*. Tübingen: Narr.
- Have, P. ten (2004). *Understanding qualitative research and ethnomethodology*. London: Sage. doi:10.4135/9780857020192
- Have, P. ten (2007). *Doing conversation analysis* (2nd ed.). London: Sage. doi:10.4135/9781849208895
- Heath, C., Hindmarsh, J., & Luff, P. (2010). *Video in qualitative research: Analysing social interaction in everyday life*. London: Sage. doi:10.4135/9781526435385
- Hirvonen, M., & Tiitula, L. (2012). Verfahren der Hörbarmachung von Raum: Analyse einer Hörfilmsequenz. In H. Hausendorf, L. Mondada, & R. Schmitt (Eds.), *Raum als interaktive Ressource* (pp. 381–427). Tübingen: Narr.
- Jääskeläinen, R. (1999). *Tapping the process: An explorative study of the cognitive and affective factors involved in translating*. Joensuu: University of Joensuu.
- Jiménez-Crespo, Miguel A. (2017). *Crowdsourcing and online collaborative translations: Expanding the limits of translation studies*. Amsterdam: John Benjamins. doi:10.1075/btl.131
- Kaindl, K. (2013). Multimodality and translation. In C. Millán & F. Bartrina (Eds.), *The Routledge handbook of translation studies* (pp. 257–269). London: Routledge.
- Kussmaul, P. (1995). *Training the translator*. Amsterdam: John Benjamins. doi:10.1075/btl.10
- Lederer, M. (2010). Interpretive approach. In *Handbook of translation studies* Volume 1 (pp. 173–179). Amsterdam: John Benjamins. doi:10.1075/hts.1.int3
- Luff, P., & Heath, C. (2015). Transcribing embodied action. In D. Tannen, H. E. Hamilton, & D. Schiffrin (Eds.), *The handbook of discourse analysis* (2nd ed.). Chichester: Wiley Blackwell.
- Luff, P., Hindmarsh, J., & Heath, C. (2000). *Workplace studies: Recovering work practice and informing system design*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511628122
- 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. doi:10.1080/08351813.2018.1413878
- Mondada, L., & Svinhufvud, K. (2016). Writing-in-interaction: Studying writing as a multimodal phenomenon in social interaction. *Language and Dialogue*, 6(1), 1–53. doi:10.1075/ld.6.1.01mon
- Muñoz-Martín, R. (2017). Processes of what models?: On the cognitive indivisibility of translation acts and events. *Translation Spaces*, 5(1), 145–161. doi:10.1075/ts.5.1.08mun

- O'Brien, S. (2011a). Introduction. In S. O'Brien (Ed.), *Cognitive explorations of translation* (pp. 1–14). IATIS Yearbook 2010. London: Continuum.
- O'Brien, S. (2011b). Collaborative translation. In *Handbook of translation studies*, Volume 2 (pp. 17–20). Amsterdam: John Benjamins. doi.org/10.1075/hts.2.col1
- Pavlović, T. (2013). The role of collaborative translation protocols (CTPs) in translation studies. *Jezikoslovlje*, 14, 549–563. Retrieved from <http://hrcak.srce.hr/file/165567>
- Risku, H., & Windhager, F. (2015). Extended translation: A sociocognitive research agenda. In M. Ehrensberger-Dow, S. Göpferich, & S. O'Brien (Eds.), *Interdisciplinarity in translation and interpreting process research* (pp. 35–47). Amsterdam: John Benjamins.
- Schegloff, E. A. (1998). Body torque. *Social Research*, 65, 535–586.
- Schegloff, E. A. (2007). *Sequence organization in interaction*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511791208
- Seiffert, A. (2005). Räumliches hören: Eine schemaorientierte Analyse der audiodeskriptiven Darstellung der Handlungsräume. In U. Fix (Ed.), *Hörfilm: Bildkompensation durch Sprache* (pp. 67–86). Berlin: Schmidt.
- Shih, C. Y. (2015). Problem-solving and decision-making in translation revision: Two case studies. *Across Languages and Cultures*, 16(1), 69–92. doi:10.1556/084.2015.16.1.4
- Sidnell, J., & Stivers, T. (2013). *The handbook of conversation analysis*. Chichester: Wiley-Blackwell.
- Walker, L. (2017). Communality and conflict: An examination of 'voice' in the Knopf translation of Mishima Yukio's *Spring Snow*. In K. Taivalkoski-Shilov, L. Tiittula, & M. Koponen (Eds.), *Communities in translation and interpreting* (pp.159–182). Montréal: Éditions québécoises de l'oeuvre.
- Warglien, M., & Gärdenfors, P. (2015). Meaning negotiation. In F. Zenker & P. Gärdenfors (Eds.), *Applications of conceptual spaces* (pp. 79–94). Cham: Springer.

¹ This study is part of the research project MUTABLE (Multimodal Translation with the Blind), and the data analysed stems from its corpus. The project is funded by the Academy of Finland. We are indebted to the informants who have participated in the research and grateful to the assistants who have helped to manage the corpus.

² The film is a German drama made for television. We do not disclose more information about the film in order to protect the privacy of the audio describers.

³ Despite numerous attempts, neither we nor our native German-speaking assistant was able to figure out this utterance from Ines. This problem illustrates the importance of technical quality in recording: if the speaker had had a microphone close to her mouth, the sound quality would have been better.